Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 37, NUMBER 2 FEBRUARY, 1947

A New One-Step Photographic Process

EDWIN H. LAND
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts
(Received January 6, 1947)

This paper describes a process which produces finished poisitive pictures, directly from the camera, in about one minute after the exposure. The camera is described as containing a strip of paper in addition to the negative material and as feeding this strip of paper, in contact with the exposed negative, through simple pressure rollers and thence out of the camera box. The process to be used with the camera is described as having but one step, instead of the many steps of conventional photography; and it is shown that this can be achieved by associating with the strips a reagent which, when spread between the two strips by the pressure of the rollers, is capable of developing the silver halide of the negative and forming the positive image at one and the same time. Several classes of processes are discussed, their characteristics explored and certain principles established for obtaining satisfactory picture quality, stability, speed of operation, etc. In one of these processes, the reagent spread between the negative and the positive strip consists of a small amount of viscous liquid containing, in rather high concentration, the necessary constituents for developing the negative image, for forming at the same time a silver complex with the unexposed grains in the negative, for transferring the soluble complex to the positive sheet, and there creating and stabilizing the positive silver image. This process runs to completion in about one minute. When the two strips are peeled apart, both are essentially dry. One strip is the finished positive picture. The process operates at temperatures from less than 30°F to over 100°F. The paper discusses the control of the rates of reduction, silver ion complex formation, and ion diffusion in the several reaction fronts; principles of stabilization of the positive picture; control of the relative rates of growth of density in negative and positive; conversion of the silver ions to particles of silver of adequate size and the control of the color of the image as a function of particle size; and various characteristic curves which have been obtained with different negative materials. These factors are then interpreted with relation to the photographic usefulness of the process.

I. INTRODUCTION

SEVERAL years ago the writer undertook the problem of devising a camera and a photographic process that would produce a finished positive print, directly from the camera, immediately after exposure. From the point of view of the user, the camera was to look essentially like an ordinary camera, the process was to be dry, the film was to be loaded in one of the usual ways, the positive print was to look essentially like a conventional paper print, and this print was to be completed within a minute or two after the picture was taken. This paper describes the development of processes of this kind and a camera to be used with them.

In photographic literature there appear to be no references to such a system of photography, even as an ideal. Recently, however, there has

been one suggestion of the need for a dry and approximately instantaneous camera.[1]

While there seems to be little record of activity leading to a completely simplified camera, much has been done to reduce the number of steps in ordinary photographic processing. A brief review of standard processing, and simplifications of it, follows:

a. Conventional Processing[2]

Expose, develop the negative, rinse, fix, wash, dry, expose the positive through the negative, develop, rinse, fix, wash, dry.

b. Reversal Processing,[3] as Used for 16-Millimeter Home Movies

Expose, develop, rinse, dissolve out developed silver, wash, expose with a predetermined flash (to give over-all exposure to remaining grains as a function of average initial exposure), develop, rinse, fix, dry.

c. Tintype or Ferrotype[4]

Expose a special emulsion (coated on a black support, historically in collodion rather than in gelatin), develop to grayish rather than black silver, rinse, fix, wash, dry. This process depends on using a black background, and obtains the whites by using the silver grains not as absorbers to provide blacks, but as diffusers and reflectors. The images seem dull by modern standards, although frequently attractive because of the delicacy of tone rendition.

d. Reversal by Solarization[5]

Expose (on an emulsion so prepared, by pre-exposure to light or by chemical treatment, that a further exposure leads to a *decrease* in density from maximum upon development), develop, rinse, fix, wash. This process is much too slow for general use. It occasionally is useful for obtaining direct positive copies in a short time, and in the studio or darkroom.

II. PRINCIPLES FOR A ONE-STEP PROCESS

In formulating the principles to be followed in developing a one-step process, certain premises were adopted concerning the sensitivity, and simplicity of the materials and mechanisms to be used.

a. The Photo-Sensitive Material

Since the ordinary silver halide negative emulsion requires development of the exposed grains to a negative image in black silver, if the maximum sensitivity of the emulsion is to be obtained, and since this implies an additional step or group of steps to obtain a positive image, it is tempting to consider using a photo-sensitive material other than a silver halide emulsion. However, examination of other photo-sensitive materials at present available proves completely disappointing, since the sensitivity of these other materials is of the order of 1/100,000th of that of ordinary photographic materials used in amateur cameras. Indeed, for tungsten light the fraction of sensitivity may be much smaller. Therefore, in spite of the fact that some of these materials, such as the diazo dye-couplers, can be processed readily, as by ammonia vapor for example, they cannot be considered except for specialized work.

As contrasted with all other photo-sensitive materials, the silver halide emulsion of conventional photography has the enormous advantage of forming a latent image in terms of a catalyst. Even in the formation of this catalyst, there appears to occur a remarkable collecting phenomenon. When the whole silver halide grain is struck by the photons of the incident light, all the photoelectrons generated throughout the grain are collected by a few sensitivity specks within it. The mass of the catalyst is multiplied by about 10^5 times when the silver halide is converted to silver by action of the developer. Thus, there are two important multiplication phenomena involved in the operation of the emulsion. Therefore, if one were tempted, even after discarding the ammonia-developed diazo

[1] V. Bush, Atlantic Mo. 176, 103 (1945) states, "Will there be dry photography? . . . When Brady made his Civil War pictures, the plate had to be wet at the time of exposure. Now it has to be wet during development instead. In the future perhaps it need not be wetted at all. . . . Often it would be advantageous to be able to snap the camera and to look at the picture immediately."

[2] L. P. Clerc, *Photography Theory and Practice* (Pitman Publishing Company, New York, 1937), Chaps. XXVI, XXVIII, XXIX, XXX, XXXI, XXXVIII.

[3] Reference 2. Chap. XXXIII.

[4] J. M. Eder, *History of Photography* (Columbia University Press, New York, 1945), fourth edition, Chapter XLVI.

[5] C. E. K. Mees, *The Theory of the Photographic Process* (The Macmillan Company, New York, 1942), Chap. VII.

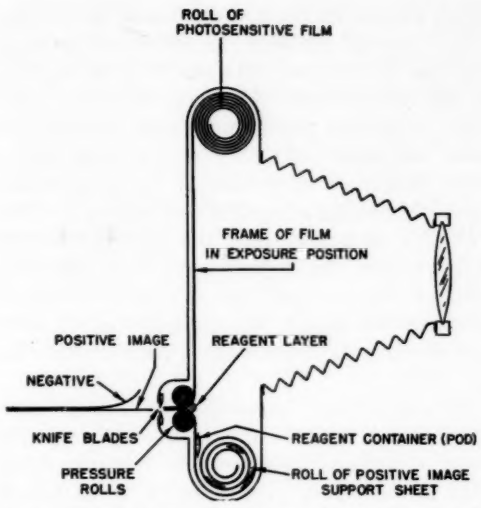

FIG. 1. Schematic diagram of camera producing one-step positive prints.

dye-coupling processes mentioned above, to venture into a research program for a new way of making an image in terms of a catalyst, it would be wise to hesitate because of the advantages of the two-stage multiplication that occurs in modern silver halide emulsions.

Finally, it is incorrect to consider photographic emulsions without also considering the reducing agents that have been found to be suitable for use with them. The modern emulsions have been designed to be used with these reducing agents, which fall in a rather limited chemical class. It is only the reducing agents in this class that can discriminate between the unexposed grains and the grains so minutely altered by exposure. Moreover, it is only when the crystals are prepared carefully and embedded in just the right kind of medium that a high degree of discrimination can be maintained by the grains, for a given developer, between the state of being reducible and the state of being non-reducible as a function only of exposure to light.

It was, therefore, taken as a ground rule for this investigation[6] that, in spite of the facts that standard emulsions yield a negative rather than a positive, and that the formation of this negative requires the use of a reducing agent, presumably associated with some amount of liquid, a way nevertheless had to be found of designing the "dry" camera to use a silver halide emulsion or some equivalent crystalline suspension.

b. The Camera

The ideal simple camera might be one in which the sensitized surface, after exposure, is subjected to some simple mechanical action, such as being passed through a pair of small steel rolls at the edge of the exposure aperture. Thus, after exposure, the film would simply be advanced through the rolls, where the pressure would develop the exposed grains to white and the unexposed grains to black. While this kind of direct development of a silver halide emulsion is, of course, fanciful, the concept of the camera itself serves as an excellent approach to the actual problems. If, for the person who used it, the camera as finally developed seemed to have the simplicity of this imaginary one, then for all practical purposes it would be as satisfactory. This imaginary camera was, therefore, taken as a symbol for the mechanics of processing.

On the two premises, then, of the use of a simple dry camera and the use of fast silver halide emulsions, the operation of the system was conceived as follows:

The camera, as shown in Fig. 1, has the usual lens, bellows, and film cartridge for the negative. An additional sheet is provided which can be in a roll in about the position where the film in an ordinary camera is wound after exposure. This second sheet feeds out of the camera, along with the exposed negative, through a pair of pressure rolls. A reagent, in the form of a minimum amount of liquid, is fed between the two sheets just before they enter the pressure rolls, so that it is spread as an extremely thin layer between the two sheets, temporarily bonding them together. The sheets have outer surfaces opaque to actinic light to protect the negative from being fogged. The thin reagent layer develops the negative, and forms the positive on one of the inner surfaces. The small amount of liquid in the reagent is imbibed into one or both of the surfaces. After a reasonable processing time, about a minute, the two sheets are peeled apart.

[6] The camera mechanisms and techniques of providing developer that will be described are useful, however, with other photo-sensitive surfaces, in those special fields where strong actinic illumination can be provided artificially.

Both are found to be essentially dry. On one of them is the finished positive image. This may be either on top of the negative or on the other sheet. The latter position has the advantage of automatically removing the geometric reversal that occurs when the positive image lies over the negative.

To meet the requirement that for the user the camera be "dry," it is desirable that no tanks, tubes, or injection systems be built into it (although they might be feasible for special models for industrial applications). Therefore, it is desirable to associate the reagent with the negative or the positive sheet. Three ways of doing this are being investigated: (1) The liquid is contained in minute, frangible, sealed cells in a layer inside the negative or the positive sheet. (2) The liquid is in a viscous plastic coating on the surface of the positive sheet which is rolled up in such a way as to seal the liquid within the roll. (3) The reagent is contained in small pods that are as wide as the sheet but only a small fraction of its length.[7]

The camera and film and reagent-spreading techniques described make it possible for the operator to expose the picture and immediately to advance from the camera the exposed frame, covered by the associated sheet. The opaque sandwich is cut or torn off and, when stripped apart a minute later, has on one of the inner surfaces the finished positive image.

Obviously, a heavy burden of requirements is being thrown on the composition of the reagent that is spread between the sheets and perhaps, also, on the dry reagents that may be included in the outer sheets. Moreover, there is another very important requirement that has not yet been discussed: reliability of operation over a wide range of temperatures. Since ordinarily chemical reactions double the velocity with every 10°C rise in temperature, the reagent must not only develop the negative and form the fixed positive, but it must also accomplish these objectives with acceptable uniformity over a serviceable range of temperatures, perhaps from 30° to 100°F.

Finally, before discussing the classification of processes in which the reagent may be called upon to perform these tasks, it should be noted that if the whole process is to be completed within a minute, it is economical of time to have the positive and negative processes occur concomitantly. Furthermore, it will be shown later that this economy of time is important in obtaining high resolving power.

III. CLASSIFICATION OF PROCESSES

So many requirements are to be made of the thin layer of reagent that it seems desirable to consider these requirements systematically. The chemical detail of the processes will be discussed in Section VI.

The general question that will now be examined is: In what ways might a reagent act to form a negative and a positive at the same time?

Class A—Exhausted Developer Processes

Assume that the reagent that is spread over the exposed silver halide emulsion layer in the negative has the characteristics of an ordinary developer, except that it is *concentrated* enough to have within its very small volume the molecules needed to complete the development of the negative and *alkaline* enough to effect prompt development. Shortly after the reagent is spread, the negative image will have developed, and in so doing, it will have formed in the reagent layer an image in terms of oxidized developing agent. The exposed silver halide grains will, as they develop, oxidize the reagent responsible for their development, and this oxidized reagent will at least for a short time remain in the neighborhood

[7] Since this third way seems most promising, some details follow: The pod is a water- and oxygen-opaque sheet, folded, sealed at the ends, filled with reagent, and bonded along the long edge. This edge-bond is so made that when the pod passes through the rolls, the hydraulic pressure of the liquid within bursts this edge and the small quantity of liquid is then transported ahead of the rolls and between the two sheets. The quantity of liquid is so metered that there is negligible excess over that required for making the very thin layer. Provision is made for the small excess so that its existence is not apparent to the operator. The pods may be fastened either to the roll of negative material or to the roll of positive material so that they are automatically carried into place as the two sheets feed through the rolls. One extremely important detail is that if the reagent were to have the viscosity of ordinary developing solutions, rather complicated devices would be required to spread it uniformly. But as the reagent is made more viscous, the spreading problem becomes much simpler. An interesting consequence of the use of viscous reagent is that after the small amount of water in the reagent has been imbibed in the adjacent layers, there is left on one of them the very thin film of the material that was used to render the reagent viscous. In some of the processes to be discussed, the positive image is formed within this thin layer.

of the exposed grains. If the positive layer is an ordinary photographic emulsion that has been pre-exposed *uniformly*, then it, too, will develop; but the amount of development at any point will be limited by the amount of *unoxidized* developing agent available at that point. There will thus be formed on this ordinary photographic paper a positive image. This image will be unsatisfactory for our purposes in at least two of three ways: (1) it will not be fixed; (2) either the highlights will not be white, because development is allowed to start in them immediately upon the spreading of the reagent and before the development of the negative has proceeded far enough to exhaust the developing agent appreciably; or (3) the definition of the positive image will be impaired because development of the positive is delayed (to preserve contrast) so that lateral diffusion of the developing agent occurs during the time of development of the negative.

Thus, this simplest of approaches involving the simultaneous development of negative and positive would at first appear not to be suitable. (Actually, in the primary process to be discussed in Class C, the phenomenon of developer exhaustion is made use of, not so much for formation of the image, as for assistance in stabilization of the image.) Some devices can be employed to improve developer exhaustion techniques. (1) A barrier layer of plastic, coated over the positive sheet, can be used to delay development of the positive until development of the negative is complete. (2) The developing agent, in an inert state, may be incorporated in the emulsion layer, so that development of the positive is not initiated until exhaustion of the developer is accomplished.

Class B—Oxidized Developer Processes

These processes, which have been explored with some success, call for forming from the oxidized developing agent either a bleach for a black dye, or a white pigment to be precipitated on a black surface. In spite of lack of experience with the questions of photographic quality involved in forming a white image on a black background, this class of processes was examined seriously because of the simple quantitative relationship that can exist between the number of grains developed, the number of molecules of oxidized developing agent formed, and the number of molecules of white pigment resulting. The pigment is produced by the reaction of the oxidized developing agent with additional reagent located either in the viscous layer or in one of the outer sheets of the sandwich.

Loss of contrast does not occur in Class B because the positive cannot start forming until a reaction has occurred in the negative. Loss of definition can be minimized because it is possible to utilize the reactions of development as fast as they occur to provide the material for the positive image. This class thus satisfies a necessary condition for high resolving power.

Class C—Soluble Silver Complex Processes

It is with the process now to be described that most of the remainder of this paper (after Class D) will be concerned. It is characterized by the simultaneous development of positive and negative images in adjacent layers, the positive image being formed from the silver halide of the unexposed grains of the negative. In the ordinary reversal process used for amateur motion picture film described above, it will be recalled that the positive is made by first developing the exposed grains, removing the silver, then fogging and developing the remaining grains. That is, stated generally, the positive is made from the unexposed grains in the negative. In this general sense, the processes of Class C bear a resemblance to the ordinary reversal process. In the Class C processes, the reagent dissolves the unexposed grains while it is developing the exposed grains and precipitates the solution of unexposed grains as a metallic silver positive in a layer adjacent to the negative. The reagent is spread between the negative and positive sheets, diffuses into the negative layer, develops it, and at the same time removes from it the unexposed grains which it then reduces to metallic silver, either in the thin layer of reagent or on the surface of the positive sheet, so that when the negative is stripped away the positive image remains. The positive image thus formed is similar to that of an ordinary print in that the highlights are the regions where no silver is precipitated and the shadows are the regions where varying amounts of silver have been precipitated.

These processes were developed in response to the fourfold question: Is it possible to form a reagent which will (1) develop to silver those grains in the emulsion that would be developed in an ordinary tray or tank developer, and (2) at the same time dissolve the other grains and yet (3) allow this solution of unexposed grains to diffuse out of the negative layer, and also (4) cause the metal ion in the solution to precipitate as the positive image in the layer adjacent to the photo-sensitive emulsion? In view of the problems previously discussed, some additional questions must be asked about a reagent for Class C: Can this reagent be so made that no image can form in the positive until the negative image has begun to be developed? Can the image in dissolved halide be utilized as fast as it is formed so that time for lateral diffusion is minimized? Can the whole process of simultaneous formation of negative and positive proceed rapidly enough to complete the picture within a minute? Finally, can the relative rates of dissolving, diffusing, and precipitating on the one hand, and developing on the other hand, be kept sufficiently invariant with temperature so that pictures of substantially the same quality will be formed over a wide range of temperatures?

There is some limited experimental support for the feasibility of a reagent which will combine the functions of developing the exposed grains and dissolving the unexposed grains.[8] In these experiments, it was found possible to make up preparations of developers containing hypo which would, for certain emulsions and at certain temperatures, develop and fix a negative in one bath over a period of about twenty minutes.[9] For one-step processes, the problems to be solved are far more difficult. It is required not only that the process work satisfactorily over a relatively wide temperature range, but also that the processing time be reduced to a minute, and that a usefully sharp and dense image be formed from the solution of unexposed grains.

Now, if there existed a solvent that discriminated directly and effectively between unexposed and exposed grains in favor of the former, the transport of the silver ions would be greatly simplified. All that would be required would be to subject the negative to the solvent action, and to precipitate the silver in a layer outside the negative. However, this kind of solvent would be a radically new kind of developer since, for it, the latent image would be a catalyst inhibiting solution. Development would mean that the reagent would ignore the exposed grains and dissolve the unexposed. Ordinary hypo and other silver halide solvents do discriminate very slightly between exposed and unexposed grains in a silver halide emulsion, the unexposed grains having a slightly greater solubility than the exposed. The discriminating effect, however, is so trivial that it cannot be utilized practically.

It is apparent, then, that since there is no satisfactory reagent for dissolving unexposed grains without dissolving exposed grains, an image-forming technique that is to utilize *unexposed* grains in solution must depend on removing exposed grains from the realm of solubility. This suggests that as the proposed reagent diffuses into the negative layer, the exposed grains must be developed into silver at least as rapidly as the unexposed grains go into solution. This dependency implies that the silver grains are essentially insoluble in the reagent that is chosen as solvent for the unexposed silver halide grains. It also implies that the unexposed grains are adequately insensitive to the reducing action of the reagent. One can thus visualize, as the reagent permeates stratum after stratum of the negative layer, an advancing reaction front in which exposed grains are not dissolving but are developing to silver and unexposed grains are not developing but are dissolving.

These are stringent demands indeed, because the requirement that the process be complete in one minute means that the developing potential of the reagent must be high. Furthermore, the requirement that the total quantity of reagent for development of negative and positive be contained in the very thin viscous layer, that is spread between the two sheets, means that the concentration of reducing agent must be extremely high. Thus the solution of silver ions must exist unreduced in the negative in a vigorous reducing environment. Similarly, the exposed but undeveloped grains must not dissolve in a vigorous dissolving environment.

[8] W. D. Richmond, "Concurrent development and fixation," Brit. J. Phot. **36**, 827–8 (1889).

[9] A. Lumiere, L. Lumiere, and A. Seyewetz, "Combined development and fixation," Brit. J. Phot. **72**, 44–45 (1925).

There is a happy combination of requirements, however, in the need (1) for having the reaction proceed rapidly, and (2) for utilizing the image in dissolved grains as rapidly as it forms. It is desirable to complete the process rapidly, not only for convenience but also because it would be extremely difficult to maintain the dissolved ions unreduced in the negative layer for a long time. Thus rapid processing makes more silver available to the positive and, incidentally, helps to limit the time of diffusion.

The processes of Class C, then, call for symmetrically advancing fronts of reaction in negative and positive, the front in the negative comprising developing and dissolving, the front in the positive comprising developing the dissolved silver ions.

One way of looking at Class C is that it provides a mechanism for developing both the exposed and the unexposed grains at the same time but in different strata. Section V(e) of the paper will discuss in detail the relative rate of growth of these two images, which are formed at about the same time.

Class D—Coupler Processes

Modern three-color photography depends on reacting the oxidized developer in place with couplers to give dyes. In some of these processes, a soluble coupler is used in the reagent, and in others the coupler is incorporated in the emulsion, either in insoluble form or associated with some insoluble organic material. Couplers can be used in one-step processes in ways analogous to Classes A, B, and C above.

Coupler Analog of Class A

If the coupler is included with the reagent so that a dye forms where the silver particles develop, the residual unoxidized developer with the coupler dissolved in it can be made to couple by an oxidizing agent at the surface of the positive sheet. This process has the same limitations as Class A (developer exhaustion) and can be similarly improved by the use of a barrier layer of plastic over the oxidizing agent, although again at a sacrifice of definition.

Coupler Analog of Class B

In the ordinary photographic color processes, the oxidized developer becomes one of the participants in the coupler reaction to form the color. These color processes yield a negative and hence are usually used in the second development stage of a reversal process. In Class B processes, the positive is obtained directly rather than after reversal by causing the coupler product to be a highlight material rather than a shadow material. A dye cannot be used but a bleach or pigment can be used in this way.

Coupler Analog of Class C

The coupler analog of Class C requires an equilibrium for the solubility of the coupler similar to the equilibrium that has been achieved in Class C for the relative solubility of exposed and unexposed grains. A coupler is present in the negative layer and its molecules have moderate mobility. As the developing agent diffuses into the negative layer, the exposed grains oxidize the developer in their neighborhood, and the oxidized developer couples with the coupler to give a dye of lower mobility. In the region of the unexposed grains, the coupler dissolves in the developer and diffuses out of the negative into the adjacent layer. If in the positive layer there is present an oxidizing agent of low mobility, then when the coupler diffusing from the unexposed portion of the negative reaches the oxidizing region, it will couple with the oxidized developer in that region. This will produce a positive image, and the color of the image will depend on the combination of developing agent and coupler that is chosen. These combinations have, of course, been thoroughly studied in color photography techniques and need not be given here. Not so much has been done to find products that will couple to give a black directly, but mixtures of couplers or developers can be used. The requirements with relation to solubility are similar to those for the silver ions in Class C. Development should just precede the dissolving reaction, and the reaction in the positive layer with the transferred solution should be rapid and should utilize the solution from each stratum as fast as the development of the negative in that stratum occurs.

IV. THE PROBLEM OF STABILIZATION

Obviously, one of the characteristics of any process which we select must be that it yields a

print which requires no washing. Since any incompleted reactions are likely to go to completion, and since reagents subject to aerial oxidation are likely to oxidize with time, particularly if the oxidation is accelerated by radiation, an unwashed print in general can be expected to become stained. Furthermore the density of the image-forming material may be destroyed by bleaching reactions if the development residues are left at random in the print.

The stabilization of the one-step positive print must thus be made an integral part of the process. In general, as much of the reagent as possible, insofar as it is not part of the image itself, should be left in the negative layer, which is stripped away. The reagents that must remain in association with the image should be left in an environment in which they do not discolor. Precautions must be taken to keep out of the final image reagents that might react to bleach the image-forming material. Particularly important among the techniques that have been used to obtain satisfactory stabilization is the use in the positive layer of components that diffuse into the reagent during the latter part of the period of image formation, to decrease the reduction potential by automatic neutralization or acidification of the reagent layer after the image is formed. Other techniques involve the use of anti-oxidants, the concentration of the excess reagent and its oxidation products in a layer other than that of the positive image, and the use of developing agents that do not form colored oxidation products.

V. DETAILED DESCRIPTION OF A CLASS C PROCESS

a. Resumé of Theories of the Nature of the Latent Image in Silver Halide Emulsions

Our experiments have led to the development of a Class C process which permits the direct production of positive images of satisfactory photographic quality. In this method, both the negative and positive images are formed from materials initially present in the negative layer; consequently, a consideration of the structure and composition of the emulsion, and the nature of the latent image was very important in the development of the process.

The silver halide emulsion[10] for a fast negative material comprises a suspension in gelatin of crystals ranging in size from 1 to 3½ microns. The crystals are essentially silver bromide and may have some iodide ions in the lattice. The observed increase in sensitivity when iodide ions are present is possibly associated with the fact that silver iodide is not isomorphous with silver bromide. In a given emulsion a group of large grains is more sensitive than a group of small grains.

The role of the gelatin is vital and complicated (it will be seen that it takes on in the Class C process duties similar to but beyond those that it has in the usual process). First, the gelatin provides for the crystals a mechanical suspension through which the components of the developing reagent, including water, can migrate readily without seriously altering the mechanical structure. Gelatin is one of the best protective colloids, and in this capacity it plays a part in the initial preference of the developer for reacting with the more exposed grains. Prior to the isolation and identification of the organic sulphur compounds in gelatin, this protein was the only reliable source of the sulfur whose presence is necessary for the preparation of sensitive negative material. Furthermore, gelatin must accept bromine atoms released in exposure, and in the course of development, it must allow bromide ions of the crystal to go into the solution while the silver ions are reduced to silver at the site of the original halide grain. Insofar as grains with low exposure tend to dissolve in the developer, the gelatin must inhibit the reduction of their silver ions until these ions can be washed out.

Each crystal possesses certain active centers or sensitivity specks. These specks are believed to consist of minute aggregates of silver sulfide whose sulfur is derived from the gelatin medium, or may be added artificially. For optimum sensitivity, there should be a limited but sufficient number of sensitivity specks in each crystal. In ordinary development, the sensitivity specks at the surface of the grain appear to be the important ones. During exposure photons strike the whole presented area of a grain and those

[10] C. E. K. Mees, *The Theory of the Photographic Process* (The Macmillan Company, New York, 1942), Part I. W. F. Berg, Ann. Rep. Chem. Soc. London **39**, 49 (1942).

that are absorbed produce within the crystal photoelectrons which are free to diffuse because their energy lies in the conductivity bands. Because of the lower level of the conductivity band of the sensitivity specks, the electrons enter them and charge them negatively. This negative charge precipitates on the sensitivity speck, as silver, a number of silver ions that are free in the crystal lattice. The resulting minute deposit of silver at the surface of the crystal constitutes a break in the ring of charged ions surrounding the whole crystal. During development a reducing agent delivers at this point a stream of electrons. Each of the latter serves for the discharge of a further silver ion of the crystal. The process results in the extrusion of microfilamentary silver and continues until the reduction of the silver halide crystal is complete, provided that sufficient developing reagent is available. The formation of the catalytic silver speck seems to occur with a quantum efficiency of one. About 200 quanta seem to be required to expose a grain.

b. The Reagent

It will be recalled that in Class C processes, a reagent is necessary which will develop the more exposed grains while at the same time it dissolves the less exposed grains. With the above picture of the nature of negative material and of the latent image in mind, let us examine in detail the requirements for such a reagent. (1) It must be a reducing agent which will develop the exposed grains very rapidly while dissolving the less exposed grains. (2) It must not readily precipitate silver ions in solution as silver in the negative, in spite of the congestion of the gelatin with freshly growing grains of silver derived from the more exposed silver halide crystals. (3) It should ultimately reduce these ions to silver either in the reagent layer or on the surface of the positive sheet.

We have found that a combination of five substances[11] satisfies these conditions: hydroquinone in the presence of sodium hydroxide serves as the reducing agent, sodium thiosulfate as the silver solvent. These reagents are dissolved in a solution of a high molecular weight material such as sodium carboxymethyl cellulose or hydroxyethyl cellulose. Sodium sulfite is also added. These reagents, except for the high molecular weight material, are familiar in ordinary photographic practice; the uses to which the combination is put in this process, however, are novel.

Both the actual operation of this process and the theory of the system are probably best introduced by a consideration of the function of the high molecular weight material. This plastic serves a variety of important purposes. The viscosity that the plastic imparts to the developer is valuable in the mechanics of bursting the pod, in spreading the reagent neatly between the two layers, and in insuring that they are immediately and uniformly wetted. It serves as a spacer between the two outer sheets, holding them far enough apart to provide an adequately large reservoir and sink for the ions and molecules that must enter and leave it, and yet it keeps the surfaces near enough together and is thin enough so that the image particles precipitated in it are made up from dissolved components that have arrived along a very short radius of diffusion. In some cases, this plastic interlayer serves as a support for the opaque particles of the positive. The two sheets as they emerge from the camera are bonded together just tightly enough so that they do not separate spontaneously during the time of processing and yet, when they have imbibed most of the water from the developer in the course of the time that they are bonded, they can be peeled apart. Either of them, depending upon design, may carry with it the extremely thin film of plastic.

The plastic molecules also exert an interesting protective colloid action. Several different methods for the production of positive images with this reagent have been developed. In one of them the plastic acts as a poor protective colloid compared to the gelatin of the negative, while in a second, the protective colloid action is sufficient to prevent precipitation until the ions reach the locus of the positive image.

When this viscous developing agent is spread between the negative emulsion and the positive supporting layers, the excellent wetting properties of the reagent combined with its high

[11] It will be clear from the discussion in Section III that many other combinations are possible; a large number has been investigated, but the discussion in this paper is confined to the combination which has been most thoroughly studied.

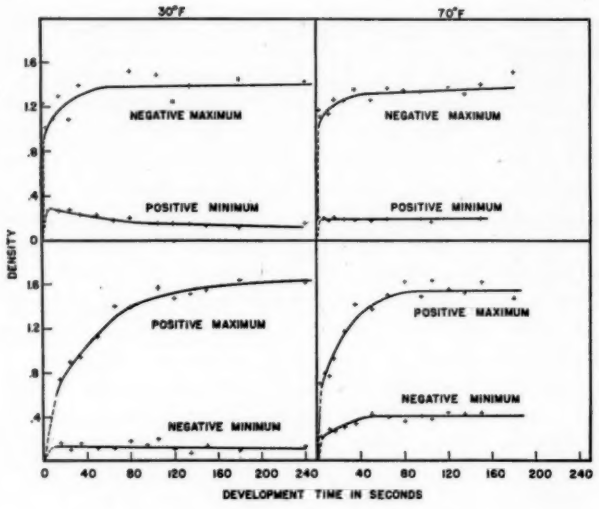

FIG. 2. Plot of growth with development time of negative and positive densities (both maximum and minimum) at 30°F and at 70°F.

development potential lead to immediate reduction of the more exposed grains. Considerable negative density is apparent almost immediately. (Cf. Figs. 2 and 3.*) A consideration of the properties of emulsions leads us to believe that the largest grains of silver halide in the surface layer of the emulsion are attacked first by the hydroquinone, while the smallest grains in the same layer are attacked by the thiosulfate ion with the formation of a soluble silver-thiosulfate complex ion. As the reaction front moves into the negative the reaction in the surface layer continues, the hydroquinone attacking successively smaller and smaller grains while the hypo continues the solution process, attacking larger and larger grains. In the end, all of these surface grains have been converted, some to silver and some to silver-thiosulfate complex ion. During the process the reaction boundary moves deeper and deeper into the negative layer, and in its wake follows the train of reactions described in detail for the surface layer.

It is clear that the ratio between the rate of formation of silver and the rate of formation of thiosulfate complexes is an important determinant of the character of the positive image. If the thiosulfate reaction is very rapid, as compared with the deposition of negative silver, only a weak negative is formed. On the other hand, when the silver forming reaction in the negative is too rapid with relation to the formation of the positive, a fogged negative and a weak positive are formed, since then the reagent is such that once it has developed the more exposed grains it may continue developing, encroaching on the grains which in the usual photographic procedure would remain undeveloped. This may happen for three reasons: (1) the reagent has a high development potential, (2) ordinary restraining agents such as sodium bromide may be omitted since they may restrain the development of the positive, and (3) there is a limit to the demand that can be made on the gelatin as a protective colloid. Therefore, it may be just as important that the development boundary in the negative not precede the solution boundary considerably, as it is that the solution boundary not precede the development.

There is a striking difference between the ways in which the hydroquinone and the hypo are used. As the reaction front advances into the negative layer the hydroquinone is taken to a higher oxidation state by each grain of silver halide which it develops. In this process no use is made of this oxidation product. The thiosulfate ion, on the other hand, which migrates in gelatin almost as rapidly as in plain water, is converted

* It should be noted that all positive curves in this paper are curves representing the entire process, from exposure of negative to development of positive, whereas the negative curves are of the usual type. Some of the negative curves represent tank development, others the new process, as indicated in the legends. All densities are measured by reflection except for the negative in Fig. 9.

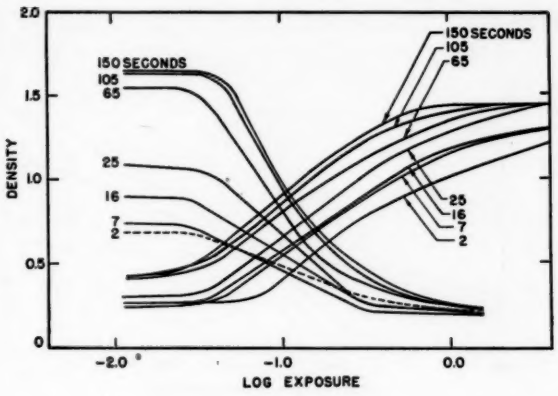

FIG. 3. Characteristic curves for one-step positive prints and their negatives at the end of development times of from 2 to 150 seconds.

to silver thiosulfate complex, which diffuses rapidly into the viscous layer where it is liberated, while the silver which it carries is deposited. Thus a small concentration of thiosulfate ions can be used in an essentially cyclic process.*

c. Mechanics of Precipitation

In the ordinary photographic negative or print the opaque grain is formed as described above by the reduction of the silver ions in the crystal *in situ*. In our process the exposed silver halide in the negative is presumably developed in the conventional way although electron microscopic studies would be desirable to determine whether the thiosulfate ions attack appreciably the grains which are to develop.

In the Class C process it is necessary to dissolve silver halide grains that are not properly members of the negative image, to transport the silver ions from these grains to an adjacent layer, and to convert these silver ions to silver in particles that are large enough to have the desired color and which are located near the normal projection (from the plane of the negative image) of their parent silver halide grains. Since we have discussed the technique of choosing the right crystal for solution and the transport of the ions, there remains then to be considered the method of forming silver atoms from the ions and the methods of aggregation of these atoms.

The silver ions are released by breaking up the silver thiosulfate complex. Three ways of doing this are: (a) by building up a concentration of complex ions in the reagent layer. When the concentration in this reducing agent becomes high enough, silver atoms are apparently formed which then condense together to form silver particles; (b) by reacting with the thiosulfate portion of the complex an ion of a metal which forms an insoluble thiosulfate; (c) by reacting with the silver ion a sulfide, selenide, or similar ion to form a speck of the silver salt. Around this speck silver particles seem to form, taking their electrons from the developing agent.** In practice processes (a), (b), and (c) can occur concurrently and are often so employed, the emphasis being shifted from one to another depending on the particular photographic problem.

(a) The concentration and condensation process. This has the advantage of directly producing silver aggregates large enough to give blue-black images. The precipitation does not occur until enough silver atoms are available and is presumably initiated by a relatively few accidental centers. These conditions are suitable for forming large precipitated particles. This process is particularly adapted

* This whole process offers singular opportunity for studying some of the mechanisms of development. The active part of the latent image in an ordinary non-solvent developer appears to be at the surface of the grain. The color sensitizing dyes are adsorbed at the surface of the grain and it seems worth investigating the local of the latent image formed through the medium of dye. Since this process uses such vigorous solvent action along with the developing agent it may be that latent image components on the inside of the grain are significant for it and that perhaps a shift in color sensitivity might be found. Some preliminary evidence has been found that there is such a shift in color sensitivity. This investigation will be reported in a later paper.

** An analogy to the theory of the development of the latent image is possible. It can be argued that the developing agent provides electrons to the silver sulfide speck, just as it is proposed that it provides electrons to the silver speck of the latent image, and that the arriving complex silver ions are discharged with precipitation of their silver on the charged silver sulfide specks.

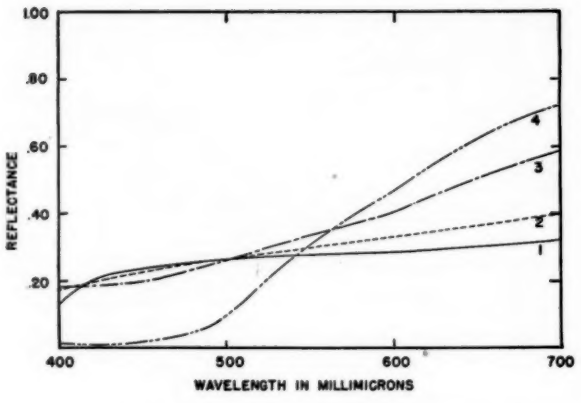

FIG. 4. Spectral reflectance curves of a range of colors of one-step positive images.

to those emulsions in which there are marked differences between the exposed and unexposed grains, and in which the contrast and gamma are high. With such emulsions the condensation process can give extremely high gamma and good blacks and whites.

In the condensation process precipitation is not related in a simple way to the availability of silver halide from the negative because precipitation in the positive waits upon some specified degree of saturation which is a function of the environment and then occurs vigorously. If the concentration is lower than this specified amount, precipitation may not occur. This is why this process is particularly suitable for those problems requiring all-or-none effects, although it can nevertheless be used for producing continuous and graded scales with the right choice of negative material.

(b) The use of metallic ions to form insoluble thiosulfates. This variant of process (a) is useful for building up the density and for making it possible to use smaller amounts of solvent. Since the precipitation of the silver again depends on the accumulation of silver atoms on a relatively few accidental centers, blue-black tones are produced.

(c) The use of anions such as the sulfide or selenide ions to produce an initial insoluble grain. These ions break up the complex by reacting directly with the silver ion. If the sulfide ions are widely distributed, as they would be if in solution, an enormous number of centers of growth is present and the silver formed on these centers is in a highly dispersed state. Images made under these conditions will be yellow or orange. Since precipitation is initiated by the sulfide centers rather than by high concentration, a relatively small amount of solvent can be used. In processes (a) and (c) when the silver is precipitated, the thiosulfate portion of the complex is left free to be used over again and this is particularly important in process (c) since the large concentration of silver (and solvent) is not required to initiate grain growth. The techniques of process (c) have proved to be particularly desirable for forming scales in the positive that have an excellent photographic relationship to the scale in the negative. For this reason it is important to consider how to obtain dark brown, brown-black, or blue-black images by this type of process.

As indicated above, an excessive number of small particles comprising the image presents a serious problem.

If the particles are only slightly enlarged by the addition of silver, the image is bright yellow. If on the other hand sufficient silver is presented for the growth of these particles to a size large enough to give a dark brown or a black image, then the density is too high in the middle tones and highlights: the resulting image is dark and has very low contrast. This difficulty cannot be surmounted by the use of a smaller number of sulfide ions. We have computed that if black images of good contrast are to be obtained, only 10^{11} ions per cubic centimeter of reagent are permissible. This is equivalent to a concentration of one part in ten billion, which is of course absurdly low.

The next step was to put on the surface of the positive support sheet, microscopic crystals of a metallic sulfide in very low concentration. The composition of these crystals was so chosen that they would dissolve slowly in the viscous reagent. It was found possible to prepare crystals whose rate of solution (or the mobility of whose dissolved ions) was low enough so that sulfide ions remained in the environment of the minute crystals until silver ions from the negative reached this location. Consequently the cluster of sulfide ions precipitated a cluster of silver sulfide particles, close enough together so that additional silver precipitated on them aggregated to masses substantially larger than those produced by dissolved sulfide ions as described above.

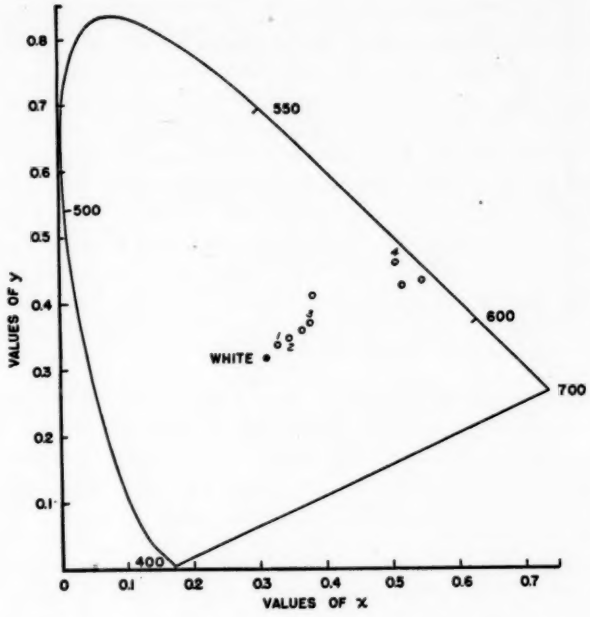

FIG. 5. Distribution on color triangle of a range of colors of one-step positive images (numbered circles correspond to curves on spectral reflectance chart).

Another fundamental advance made it possible to obtain still larger masses of silver. Techniques which will be described in a subsequent paper were developed for forming low density galaxies of extremely small sulfide crystals loosely clustered together. Although these galaxies were in some cases made of sulfides with a high optical absorption coefficient, the mass density of the galaxy was so low that little or no optical density was apparent when the positive sheet was covered with these groups of microscopic crystals. In this variant of the process, when the viscous reagent was spread over the sheet covered with these galaxies, very low density clouds of sulfide ions were available, the diameter of which was was of the order of the diameter desired for the silver particles which composed the final image. When the silver complex ions arrived at these galaxies, silver sulfide was formed at each of the sulfide ions and was covered immediately by a mass of metallic silver reduced from the other ions in solution. In this way dark brown, brown-black, and blue-black images were obtained.

d. Color of the Image

It is an elementary fact of colloidal chemistry that metallic sols are colored. The range of size for particles that can stay in suspension in the ordinary sol is of a lower order than that involved in making good neutral photographs. The images described in the preceding section, made by initiating precipitation with a trace of sodium sulfide dissolved in the viscous reagent, are yellow and are much the same color as a classical silver sol.

A careful study of the change in color of silver images was made by Chapman Jones[12] by depositing known quantities of mercury on small particles in the silver image. His approach was pragmatic and accurate, and produced a whole range of colors in photographic images. Nevertheless, his problem was different from that of this paper. He used in his experiments an already existing silver image, and deposited mercury on the particles that composed the image. The problem presented by the new process is to grow particles in the correct place directly from solution and to grow them, as nearly as possible, all to the same size. The importance of the ion galaxy techniques discussed in the preceding section is that with those techniques the particle size of the silver seems to be determined by the diameter of the galaxy. The silver aggregates act as if there were metallic conduction between the particles early

[12] C. Jones, Phot. J. (RPS) 51, 159–174 (1911).

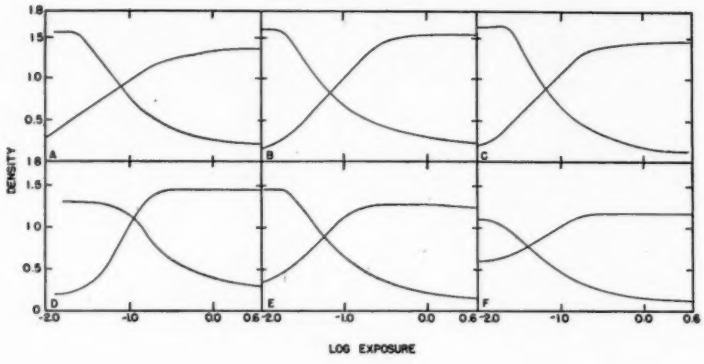

FIG. 6. Characteristic curves of a variety of tank-developed negatives and corresponding one-step positive prints.

in the course of the deposition of the silver on the galaxy, so from the start, the particle acts like a large particle. Its absorption coefficient increases as deposition proceeds, but its color seems to remain nearly constant with a slight change from brown towards black. It is scientifically interesting and useful in practice to be able to make images over the wide range of colors that can be had by developing the particles in the positive image to different sizes in different pictures.

Figure 4 shows the spectral reflectance curves of a range of color that can be obtained with this process, and Fig. 5 shows the position of the colors of the same samples on the color triangle. The samples were chosen to have a reflection density of 0.5 approximately in the middle of the visible spectrum. The trend from the side to the middle of the triangle shows the progression of image color toward a neutral gray as the sulfide ions are provided, first, in solution, second, as small crystals uniformly distributed on the positive image sheet, and then, as galaxies of *very* small crystals.[13]

e. Review of the Simultaneous Growth of Negative and Positive

Now that the fate of the individual particles has been described from the time of spreading of the reagent to the formation of the complete negative and positive images, the over-all growth of the positive and negative photographic images will be discussed.

Figure 3 shows a group of negative curves and the corresponding group of positive curves, the pairs being identical except for the length of the development period. For example, the right-hand number 2 points to the characteristic curve for the negative that goes with the left-hand number 2, which points to the positive characteristic curve. The value 2 indicates that the sheets were peeled apart and that image growth was stopped at the end of two seconds.

It will be noted that density has formed in the negative rapidly, that it leads the growth of density in the positive slightly, that after 30 seconds the maximum density of the positive is starting to exceed the maximum density of the negative, and that finally, at the end of the process the maximum density of the positive is considerably higher than that of the negative. This latter remarkable phenomenon might be thought to be a result of incomplete development of the negative in this process. As can be seen in Fig. 6, this is not the reason. In Fig. 6, the characteristic curves of a variety of negative materials developed to their maximum contrast by tank development are shown plotted with the characteristic curves for the positives made by the new process from negative material of the same type.

In most of the examples of Fig. 6 the maximum positive density is greater than the maximum negative density obtained by tank development. (The exceptions are usually caused by some special condition in the negative such as deliberate fogging.) The higher density of the positives means that the material of the positive has a

[13] One precaution should be observed in relating these colors to those of silver sols or of lantern slides made to show the color of different size silver particles. These are reflectance curves and may be complicated by the relation of transmitted, reflected, and back-scattered light. Studies are being started now on the transmission color alone of these suspensions and will be reported later.

higher photometric coefficient than the ordinary developed silver halide grain.[14]

Figure 2 shows negative maxima and positive minima and negative minima and positive maxima as a function of time at two different temperatures. This figure demonstrates that a satisfactory ratio obtains in the advancing reaction front between the rate of development and the rate of solution, and that this ratio is substantially maintained over a wide temperature range. The lower right-hand curves, as compared with the lower left-hand curves, indicate that the rise in temperature causes more of the less sensitive grains to develop in the negative, either directly or after dissolving. The density attributable to this silver precipitated in the negative is seen to be missing, to some extent, from the positive maximum curve just above it, as compared with the curves in the lower left.

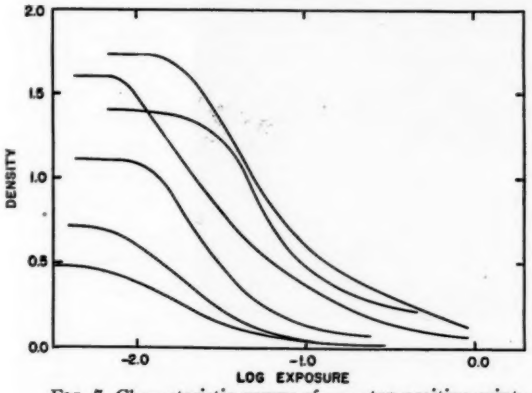

FIG. 7. Characteristic curves of one-step positive prints, made with identical negative emulsions, but with variations in treatment of the positive sheet.

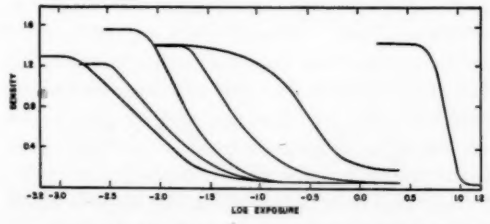

FIG. 8. Characteristic curves of one-step positive prints from a variety of negative emulsions processed against identical positive sheets.

[14] C. E. K. Mees, *The Theory of the Photographic Process*, p. 231.

With a given negative material a wide range of characteristic curves and contrasts can be obtained, depending on the relative rate of growth of the negative and positive, and on the covering power of the particles. Figure 7 shows a variety of characteristic curves obtained with a single negative material by changing the type of anion clusters in the positive supporting sheet. There is, however, a limit to the variation that can be produced by the variation in the processing of the negative material, and a much wider range of characteristic curves is obtained by choice of a variety of negative emulsions with widely different characteristics. Figure 6 is an example of a set of rather similar negative materials with their corresponding positive curves. Figure 8 is a set of characteristic curves for the over-all process using a variety of very different negative emulsions. The striking differences in spread and curve shape in the latter case are obvious.

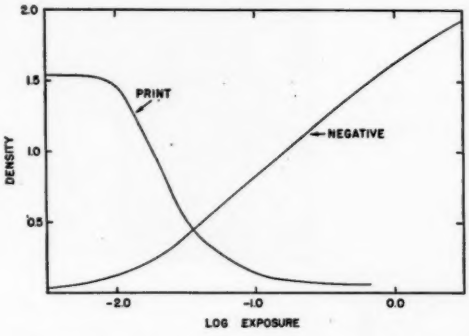

FIG. 9. Characteristic curves for a tank-developed Verichrome negative, and a one-step positive print made with an identical negative emulsion.

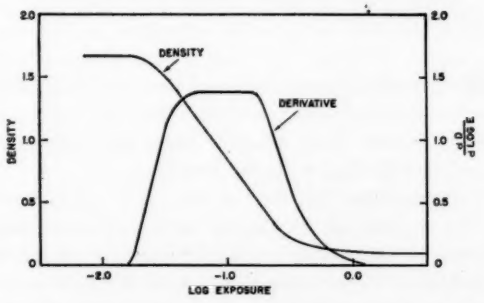

FIG. 10. Characteristic curve and its derivative curve for a one-step positive print.

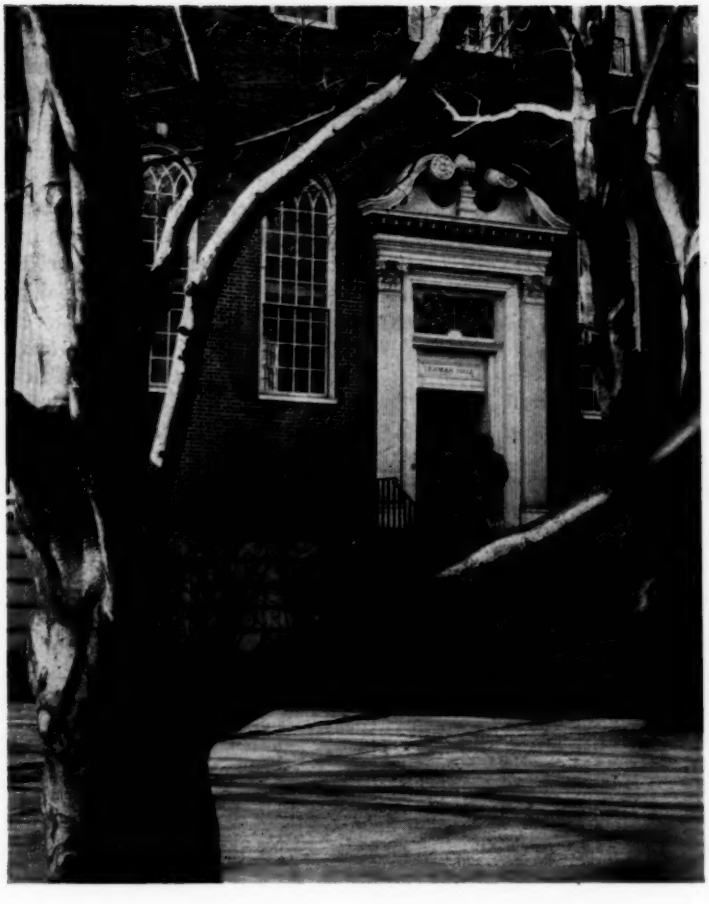

FIG. 11. A halftone made from a print made by the process.

VI. PHOTOGRAPHIC USEFULNESS OF THE ONE-STEP PROCESS

Some of the primary factors in determining the photographic usefulness of a process are the sensitivity, latitude, resolving power, shape of the characteristic curve, the range of densities, minimum density, color, and stability. Since the last four factors have already been discussed this section will deal with the first four.

The process described in this paper can be used with a range of emulsions of varied sensitivity from contact paper to highly sensitive negative materials. Frequently the process appears to increase speed of the negative material relative to the rated speed. This effective increase in speed can be understood from Fig. 9, in which the characteristic curve for Verichrome emulsion developed in a tank is contrasted with a positive print made from the same type of emulsion. It will be noted that the characteristic curve for the positive lies in the low exposure end of the negative curve, and that the positive process has had the effect of giving a greatly increased slope to the curve for the new process for a region of the negative curve in which the slope is small. In conventional practice, with this particular negative material, it would have been possible to expose so as to use nearly the same portion of the negative curve that the new process does, and a good print could have been made by selecting the

right paper. But the recommended speed for the negative material would have led to a greater exposure to enable the operator to take advantage of the latitude of this negative material. An "error" in exposure made by the operator would simply have shifted the significant portion of the picture along the long sloping curve of the negative response. In making the print, this "error" would have been compensated by selecting a proper paper and exposure. In the new process the operator would be required to use the smaller exposure and to treat the film as if it had only the higher sensitivity.

The latitude of any one-step process will obviously be less than for the combination of the negative material and a choice of positive papers. In this respect the new process is like most of the color photography processes. It is necessary for the best results to choose the correct exposure for the particular characteristic curve. A correct exposure lies within a range of about two stops. This limitation is perhaps offset by the fact that the operator can see the picture as taken and correct for any error at once in a new exposure.

The resolving power of the process has been found to vary between 30 lines per mm and 10 lines per mm, depending on the type of reagent, emulsion, or positive sheet that is chosen. The resolving power requirement varies with the plan for use of the image. If a large hand-held print is produced directly from the camera so that enlargement is not necessary, then images appear reasonably satisfactory with a resolving power of 10 lines per mm. If images are to be projected, a resolving power of at least 30 is desirable.

As Fig. 8 shows, a wide variety of characteristic curves can be obtained by varying the characteristics of the emulsion. This figure contains the curves obtained by using a range of commercial emulsions. Figure. 7 shows how the characteristic curve was changed by using the same emulsion but varying the composition of the reagent in the positive sheet. It is also possible to produce a somewhat similar range of curves by varying the composition of the developing reagent. Figure 10 shows a typical "good" characteristic curve obtained by using an emulsion especially prepared for the process. The curve for the derivative is plotted for the convenience of those who wish to use Jones'[15] criteria for estimating the relation of characteristic curve to the quality of the print. In general, it would appear that since a wide variety of curves can be made and since the characteristic curve of the negative can be related to the characteristic curve of the positive, any desired curve can be planned on in defining the process. A great many photographs have been made, one of which is reproduced here (Fig. 11).

ACKNOWLEDGMENT

The writer wishes to thank all of those members of the Polaroid Research Department who have participated in this investigation for their enthusiastic and effective assistance, and to note especially the contribution of his first assistant on this program, Mrs. Eudoxia Muller Woodward. The writer also wishes to thank the Eastman Kodak Company for its cooperation in supplying a variety of speciál emulsions for this investigation.

[15] Reference 14, p. 824.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 2 FEBRUARY, 1947

Brightness and Polarization of the Daylight Sky at Various Altitudes above Sea Level

R. TOUSEY AND E. O. HULBURT
Naval Research Laboratory, Washington, D. C.
(Received October 31, 1946)

INTRODUCTION

IN the following pages is described an experimental and theoretical investigation of the light of the daylight sky. The brightness and polarization of the sky at various points around the compass from the horizon to the zenith for various altitudes of the sun were measured during eleven flights in an airplane from sea level to 10,000 feet and one flight to 20,000 feet. The weather was selected, not always with success, to be clear and free from clouds. The earth reflectivity and the solar illumination were also measured.

A quantitative theory of the sky brightness and polarization was developed on the assumption that the atmosphere was composed of pure air molecules with scattering according to Rayleigh. For such an atmosphere the attenuation β of sunlight viewed with the light adapted eye was 0.0126 km^{-1}. The "U. S. standard atmosphere," the reflectivity of the earth, the polarization defect of air, the absorption of atmospheric ozone, and the solar illumination entered into the theory.

It was found that the sky data at 10,000 feet for the five flights when the air was the clearest agreed well with each other and disagreed consistently with the theoretical calculations for pure air for which $\beta=0.0126$ km^{-1}. They agreed with the theory within experimental error when β was increased to 0.017 km^{-1}. This indicated that in these five clearest cases the air above the observer was not pure air, but was air which contained impurities in the form of scattering particles small with respect to the wave-length of visible light sufficient to increase the theoretical scattering of pure air by about 35 percent. This was again found to be the case in the one flight to 20,000 feet. The data of the other seven flights indicated less clear air than in the above five flights and required β's from about 0.018 to 0.021 km^{-1} in the theory in order to obtain agreement.

OBSERVATIONS OF SKY BRIGHTNESS AND POLARIZATION

The sky brightness and polarization were measured with a Macbeth illuminometer modified slightly for the purpose. The field of view was a cone of about 4° angular diameter. To obtain a color match with the sky a blue filter was placed over the comparison lamp of the illuminometer; the match was good, but rarely perfect. The sky illuminometer was in a mount clamped to the airplane with motions in altitude and azimuth. Over the front of the instrument was placed a polarizing plate and a tube with diaphragms to reduce stray light. Measurements could be made to within about 8° of the sun, but at smaller angles direct sunshine in the instrument caused errors.

The sky illuminometer was calibrated by standard procedure to read the brightness of a source of the spectral distribution of sky light in candles per square foot. The calibration was obtained by a comparison of the instrument without the blue filter against a lamp standardized by the National Bureau of Standards, together with the photometric transmission of the blue filter determined from the I.C.I. spectral luminosity curve, and the measured spectral curves of the transmission of the blue filter and the intensity of the illuminometer lamp.

In order to measure the solar illumination the direct sunlight was allowed to pass through an open port in the plane, thereby screening off nearly all of the sky light; it illuminated a diffuse white plaque held perpendicular to the rays of the sun. The brightness of the white plaque was measured with another calibrated illuminometer equipped with a yellow filter over the external field to give a color match.

All observations were made over a region within about 200 miles of Washington, D. C., during twelve airplane flights usually to altitudes of about 10,000 feet; in Flight 9 an altitude of 20,000 feet was reached. All observations were made through an open port in the plane. The

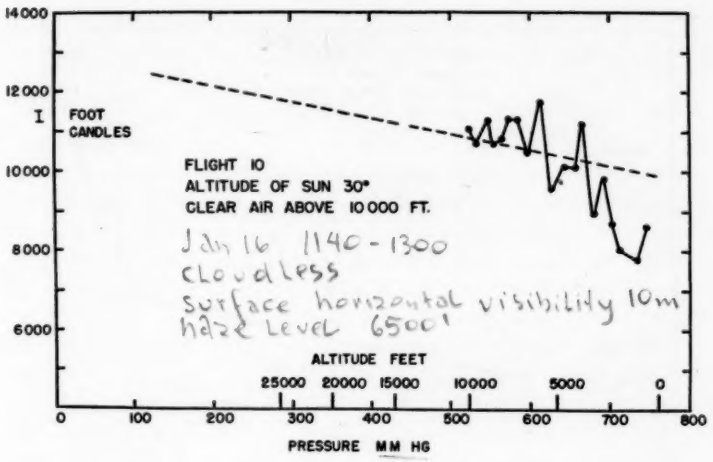

FIG. 1. Solar illumination for "clear" air above 10,000 feet.

usual program of sky observations included measurements of the zenith sky as the plane climbed from the surface to 10,000 feet, and measurements at various points in the sky at 10,000 feet. On several flights solar illumination measurements were made and special programs were carried out. The programs were modified, and often partially interrupted, by weather, clouds, and conditions of flight.

Viewed from the surface in cloudless weather with a normally clear atmosphere the region of the sky about 90° from the sun and not too near the horizon is the darkest, bluest, and most polarized region of the sky. As the angular distance from the sun decreases, the sky increases gradually in brightness and becomes whiter, and within 10 or 15 degrees of the sun the brightness and whiteness are usually very noticeable. The region near the sun is described as an "aureole" and is caused by haze of foreign particles of relatively large size. The angular extent and the brilliancy of the aureole vary with the amount and type of haze. A familiar manifestation of aureole is the glow surrounding lights on a foggy night.

At the surface around Washington, and probably in most localities, a condition of no aureole rarely occurs, indicating that the atmosphere at sea level is hardly ever free of haze. The whiteness of the sky near the horizon is usually caused by large particle haze. The smaller the size of the haze particle the greater the angular extent of the aureole, the less its whiteness and the less conspicuous it becomes. For particles small with respect to the wave-length of visible light an aureole does not appear since approximately the same type of scattering occurs as for air molecules. Such a haze can be revealed by a brightness measurement; it causes an increase in brightness but no great change in polarization, angular distribution, and blueness. Evidence of this type of haze was found at the higher altitudes of the present experiments. In brief, although the presence of an aureole is conclusive proof of haze, the fact that no aureole can be readily seen does not mean that the atmosphere consists only of air molecules.

As one ascends through a normally clear atmosphere the aureole becomes less and at a sufficient altitude may become imperceptible. It was noticed that often, but not always, one passed through most of the haze and could see the top of the haze appearing as a dark line around the horizon with relatively blue sky above. The dark line was termed the "haze level"; near this level the aureole usually, but not always, disappeared. In most of the present sky measurements the haze level and the aureole were recorded in order to give some indication of the amount of haze prevailing above the observer.

Representative observational data of four flights are plotted in Figs. 1 to 10; the data of all the flights were too numerous to publish. Since each series of measurements extended over a time interval during which the altitude of the

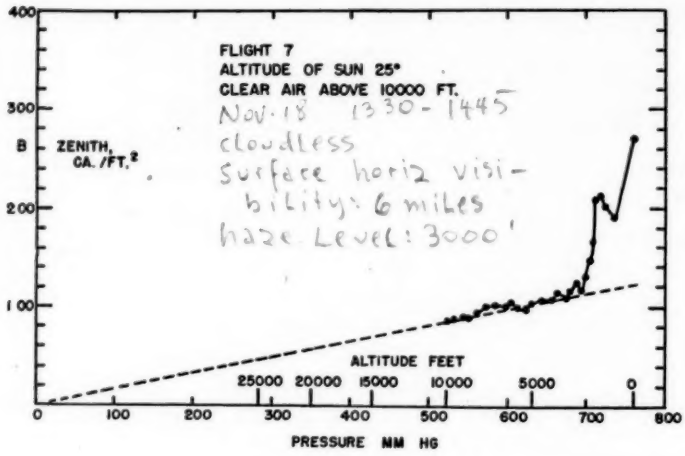

FIG. 2. Zenith sky brightness for "clear" air above 10,000 feet.

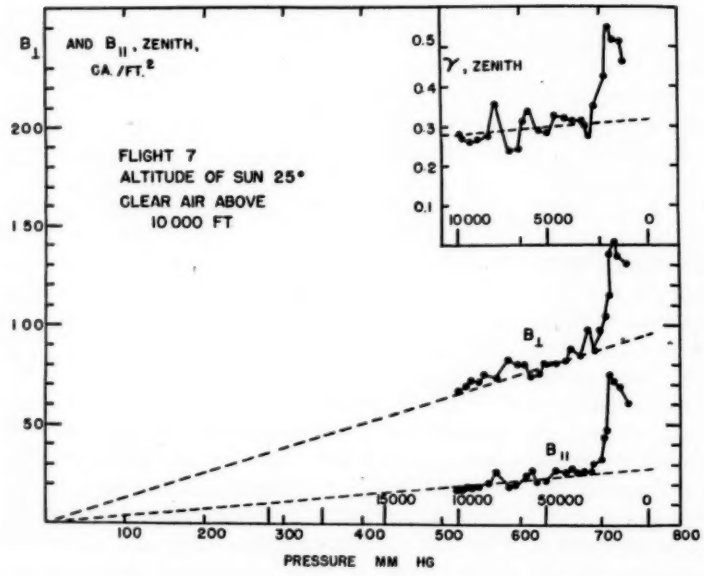

FIG. 3. Polarized components and polarization of zenith sky brightness for "clear" air above 10,000 feet.

sun changed slightly, they were corrected to refer to a single altitude by means of the slopes of the experimental or theoretical curves. The corrections were small, being usually less than 2 percent and always less than 5 percent. With these corrections the data were plotted exactly as they were taken and show evidence of experimental vicissitude due mainly to motion of the plane and lack of opportunity to repeat readings. In the figures the dots and heavy lines are the observations and the dotted curves are theoretical from Table I.

The following notation is used:

ζ zenith angle of the sun, the altitude of the sun is $90°-\zeta$;

ξ zenith angle of the point in the sky under observation, the altitude of the point above the horizon is $90°-\xi$;

Z the bearing of the point in the sky relative to the sun;

B candles per square foot, brightness of the point in the sky, $B=B_{\parallel}+B_{\perp}$;

$B_{\parallel}$ candles per square foot, the polarized component of B with electric vector in the plane passing through the sun, observer, and point in the sky;

$B_{\perp}$ candles per square foot, the polarized component of B with electric vector perpendicular to the plane;

$\gamma=B_{\parallel}/B_{\perp}$; the polarization γ ranged from 0 for complete

polarization to 1 for no polarization, and hence determined how much the sky may be darkened by a polarizing plate;

I footcandles, the solar illumination at the observer.

The data of Figs. 1 to 10 were selected to illustrate the cases of "clear" and "less clear" air above 10,000 feet; the terms are defined in the last section. "Clear" refers to skies of brightness nearest to Rayleigh theory, while "less clear" applies to skies with obvious haze. It is seen from Figs. 1 to 5, which refer to "clear" air, and from Figs. 6 to 10, which refer to "less clear" air, that I is greater and B is less for clear air than for less clear air, and γ is about the same for the two cases. The conditions of the four flights of Figs. 1 to 10 were: Flight 6, November 4, 1315 to 1430, scattered cirrus clouds, surface horizontal visibility greater than 10 miles, haze level indistinct 4000 to 7000 feet, aureole at 10,000 feet: Flight 7, November 18, 1330 to 1445, cloudless, surface horizontal visibility 6 miles, haze level 3000 feet, no aureole at 10,000 feet; Flight 10, January 16, 1140 to 1300, cloudless, surface horizontal visibility greater than 10 miles, haze level 6500 feet, faint aureole at 10,000 feet; Flight 12, February 14, 1130 to 1300, cloudless, northwest wind after four days of dry northwest winds, surface horizontal visibility 8 miles, haze level about 6000 feet but not sharply defined, wide aureole at 10,000 feet.

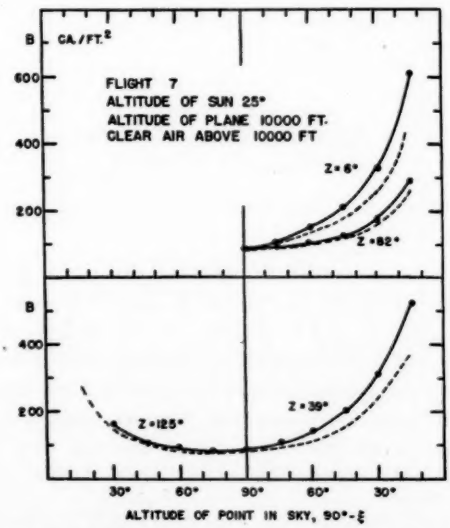

FIG. 4. Sky brightness at 10,000 feet for "clear" air.

THEORY OF SKY BRIGHTNESS AND POLARIZATION

Several theoretical treatments of sky brightness and polarization have been developed, as, for example, those of Tichanowski,[1] Pokrowski,[1] and King.[2] The developments were not adequate for application to the present measurements for they did not consider completely the effects of haze, ozone, the reflectivity of the earth, and the imperfection of polarization of the air molecules, and in no case were the theories carried through to the point of calculating the brightness and polarization of various points in the sky at various altitudes above sea level. Therefore, in the following paragraphs are presented theoretical formulas which, in some respects, are more complete than those hitherto available.

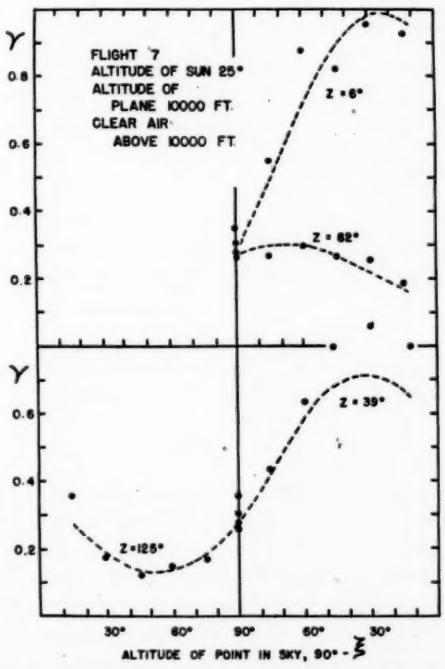

FIG. 5. Sky polarization at 10,000 feet for "clear" air.

[1] See H. Geiger and K. Scheel, *Handbuch der Physik* (1928), Vol. 19, Chapter 4.

[2] L. V. King, Phil. Trans. **212**, 375–433 (1913).

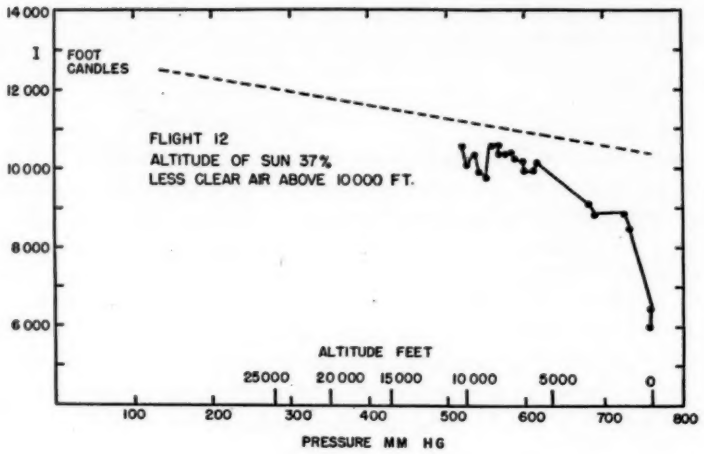

FIG. 6. Solar illumination for "less clear" air above 10,000 feet.

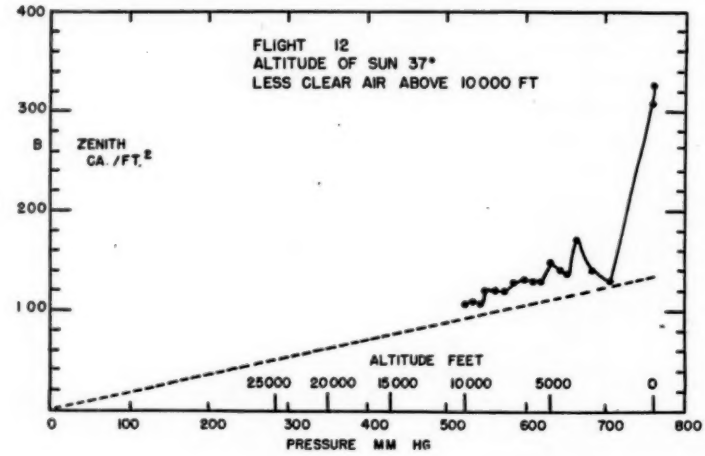

FIG. 7. Zenith sky brightness for "less clear" air above 10,000 feet.

Let B be the brightness of a point in the sky as it appears to an observer. It is assumed that

$$B = B_p + B_m. \tag{1}$$

B_p is the brightness arising from the direct rays of the sun scattered to the observer by all the air particles along his line of sight; B_p is *primary* scattered sunlight. B_m is the brightness arising from all other rays scattered to the observer by the air particles along his line of sight. Since these rays come from the surrounding atmosphere and from the surface of the earth which owe their illumination to sunlight, B_m is *multiply* scattered sunlight. We proceed to evaluate B_p and B_m.

Scattering, Polarization, and Attenuation Formulas

The optical scattering coefficient $\sigma_{\phi\lambda}$ is defined by

$$i_{\phi\lambda} = \sigma_{\phi\lambda} i_\lambda, \tag{2}$$

where i_λ is the flux per unit area of a beam of unpolarized light included in the wave-length interval λ to $\lambda + d\lambda$ which traverses unit volume of the scattering material, and $i_{\phi\lambda}$ is the flux per unit solid angle per unit volume scattered at angle ϕ to the beam. For a gas $\sigma_{\phi\lambda}$ may be derived from the Rayleigh[3] theory of the scat-

[3] Lord Rayleigh, Phil. Mag. **41**, 107 (1871); *ibid.* **47**, 375–384 (1899); Proc. Roy. Soc. **84A**, 25–46 (1911); *ibid.* **90A**, 219–225 (1914).

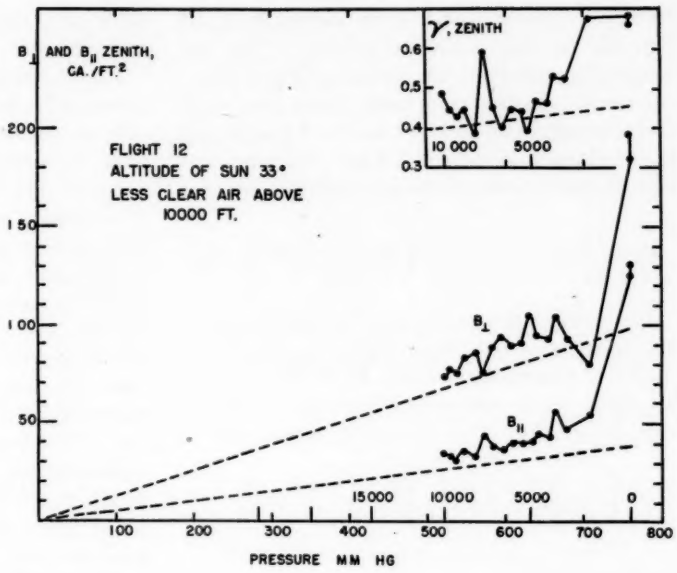

FIG. 8. Polarized components and polarization of zenith sky brightness for "less clear" air above 10,000 feet.

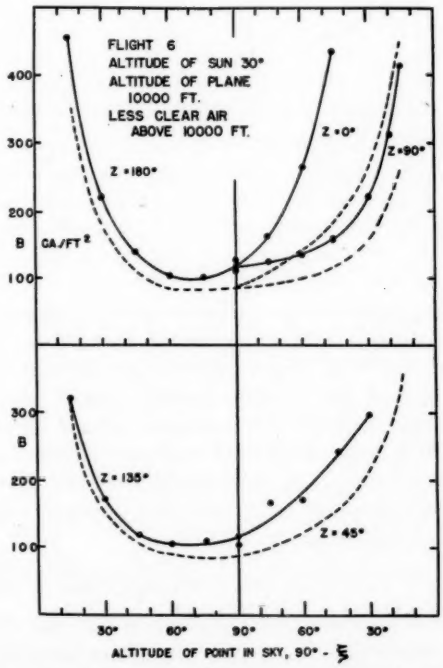

FIG. 9. Sky brightness at 10,000 feet for "less clear" air.

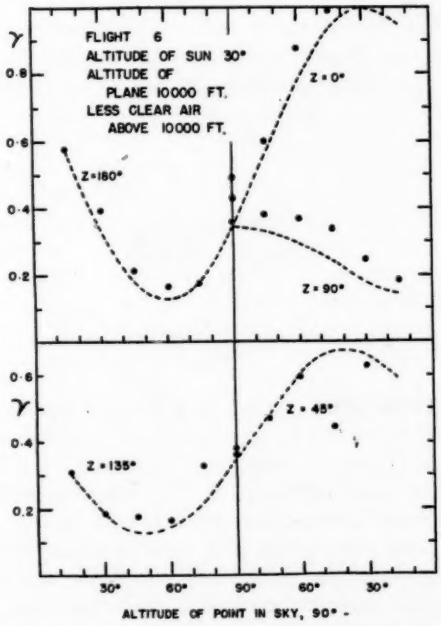

FIG. 10. Sky polarization at 10,000 feet for "less clear" air.

tering of light by particles. For a gas one finds

$$\sigma_{\phi\lambda}=\frac{2\pi^2}{n\lambda^4}(\mu_\lambda-1)^2(1+\cos^2\phi), \qquad (3)$$

where μ_λ and n are, respectively, the refractive index and the molecular density of the gas.

From the scattering theory it follows that light scattered at $\phi=\pi/2$ is completely polarized.

However, observations[4] have shown that for pure air the light scattered at $\phi=\pi/2$ was not completely polarized, the ratio δ of the weak to the strong polarized component being about 0.04 and approximately constant with λ. δ was termed the "polarization defect." When this term was introduced into the scattering theory (3) became[5]

$$\sigma_{\phi\lambda}=\frac{2\pi^2}{n\lambda^4}(\mu_\lambda-1)^2\frac{6(1+\delta)}{6-7\delta}\left(1+\frac{1-\delta}{1+\delta}\cos^2\phi\right). \quad (4)$$

In the present experiments the polarization γ has been measured. Referring to (1) B_m is assumed to be unpolarized and B_p to be polarized according to the linear oscillator theory underlying (3) and (4). Then it may be shown[5] that

$$\gamma=\left[D+\frac{B_m}{2B_P}(1+D)\right]\Big/\left[1+\frac{B_m}{2B_p}(1+D)\right], \quad (5)$$

where

$$D=\delta+(1-\delta)\cos^2\phi. \quad (6)$$

Let $B_\perp$ and $B_\parallel$ be the two polarized components of B with electric vectors perpendicular and parallel, respectively, to the plane containing the observer, the sun, and the point in the sky under observation. Then

$$B=B_\perp+B_\parallel, \quad (7)$$

and

$$B_\perp=\frac{1}{1+\gamma}B, \quad (8)$$

$$B_\parallel=\frac{\gamma}{1+\gamma}B. \quad (9)$$

Ozone exists in the atmosphere mainly in levels above 20 km, and, because of the Chappuis band, it absorbs sunlight in the visible portion of the solar spectrum. The atmosphere above the ozone amounts to less than one percent of the total atmosphere, and, therefore, the simplifying assumption is made, valid to a close approximation, that all the ozone is above the "top" of the atmosphere. As a result of the assumption we calculate the degradation of the sunlight in its passage through the ozone region and then determine the effects of the sunlight in the atmosphere.

The attenuation of visible light in the natural atmosphere below the ozone region, whether clear or rendered unclear by haze, fog, rain, and snow, is mainly caused by scattering and absorption. The only two permanent constituents of the lower atmosphere which have bands or lines of absorption in the visible spectrum are oxygen and water vapor, but estimates from available and none too complete data indicate that their attenuations due to true absorption are considerably less than those due to scattering. This is not true for certain types of smokes, but smoke is omitted from the present considerations. Therefore, it is assumed that the atmospheric attenuation for visible light arises mainly from scattering.[6]

The attenuation coefficient β_λ due to scattering is the integral of $\sigma_{\phi\lambda}$ over a sphere, or

$$\beta_\lambda=2\pi\int_0^\pi \sigma_{\phi\lambda}\sin\phi d\phi. \quad (10)$$

In the case of pure air molecules $\sigma_{\phi\lambda}$ is given by (4), and (10) becomes

$$\beta_\lambda=\frac{8\pi^3}{3}\frac{(\mu_\lambda-1)^2}{n\lambda^4}\frac{6(1+\delta)}{6-7\delta}\left(3+\frac{1-\delta}{1+\delta}\right). \quad (11)$$

Calculation of B_p

Referring to Fig. 11 let O be the position of the observer and X his zenith. Let OS be the direction of the rays of the sun and OP the line of sight of the observer. Let the zenith angles, i.e., the angles to the vertical of OS and OP, be ζ and ξ, respectively. Let the angle between OS and OP be ϕ and the projection of this in the horizontal plane be Z; Z is the bearing of the line of sight from the sun. Then

$$\cos\phi=\cos\zeta\cos\xi+\sin\zeta\sin\xi\cos Z. \quad (12)$$

It is assumed that the earth is flat and that the characteristics of the atmosphere vary only in the vertical direction. The results which we shall obtain will differ from those derived for a curved earth[7] by less than 2 percent for $\zeta=\xi$

[4] See M. Born, *Optik* (Verlagsbuchhandlung, Julius Springer, Berlin, 1933), p. 388.
[5] L. H. Dawson and E. O. Hulburt, J. Opt. Soc. Am. **31**, 554–558 (1941).
[6] E. O. Hulburt, J. Opt. Soc. Am. **25**, 125–130 (1935).
[7] E. O. Hulburt, Phys. Rev. **55**, 639–645 (1939); see page 641.

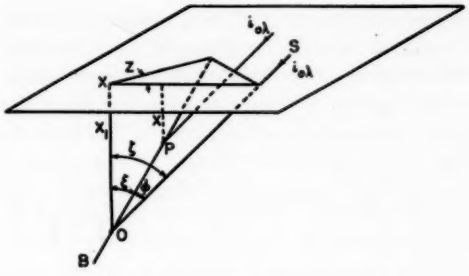

FIG. 11. Geometry of primary scattering.

$<80°$. Let x_1 be the vertical thickness of the atmosphere above the observer and x the thickness above a point P on his line of sight. The thicknesses are reduced to standard conditions at 1 atmosphere and 0°C; this introduces no error. Refraction of rays of light in the atmosphere is small, less than ½ degree, and is neglected.

Let β_λ' be the absorption coefficient of the ozone region for light of wave-length λ passing through the entire ozone region in a vertical direction. The amount of ozone in a vertical column through the region is known to vary with the year, latitude, season, and weather. But since there are no data on the amount of ozone for the times and places of this investigation, possible variances in the present calculations due to changes in ozone cannot be determined. As mentioned later the variances, although not inappreciable, are not large, probably less than 3 percent.

Let $i_{0\lambda}'$ be the intensity of sunlight outside the atmosphere for wave-length λ, and $i_{0\lambda}$ the intensity after passing through the ozone region; $i_{0\lambda}$ is the intensity which falls on the "top" of the atmosphere. Then

$$i_{0\lambda}=i_{0\lambda}'\exp(-\beta_\lambda'\sec\zeta). \tag{13}$$

The intensity reaching P, Fig. 11, is

$$i_{0\lambda}e^{-\beta_\lambda x\sec\zeta}.$$

The amount scattered toward O is this quantity multiplied by $\sigma_{\phi\lambda}$, and is attenuated by a factor

$$e^{-\beta_\lambda(x_1-x)\sec\xi}$$

in reaching the observer. The total intensity $i_{p\lambda}$ which reaches the observer from all elements along OP is

$$i_{p\lambda}=\int_0^{x_1} i_{0\lambda}\sigma_{\phi\lambda}e^{-\beta_\lambda x\sec\zeta}e^{-\beta_\lambda(x_1-x)\sec\xi}dx\sec\zeta. \tag{14}$$

Integrating

$$i_{p\lambda}=i_{0\lambda}\frac{\sigma_{\phi\lambda}}{\beta_\lambda}\frac{e^{-\beta_\lambda x_1\sec\zeta}-e^{-\beta_\lambda x_1\sec\xi}}{1-\sec\zeta\cos\xi}. \tag{15}$$

Introducing (4), (11), and (13) into (15) yields

$$i_{p\lambda}=\frac{3}{4\pi}i_{0\lambda}'\exp(-\beta_\lambda'\sec\zeta)\frac{1+\frac{1-\delta}{1+\delta}\cos^2\phi}{3+\frac{1-\delta}{1+\delta}}\times\frac{e^{-\beta_\lambda x_1\sec\zeta}-e^{-\beta_\lambda x_1\sec\xi}}{1-\sec\zeta\cos\xi}. \tag{16}$$

The brightness is $\psi_\lambda i_{p\lambda}$, where ψ_λ is the luminosity function of the light adapted eye. Then the total brightness B_p is

$$B_p=\int_0^\infty \psi_\lambda i_{p\lambda}d\lambda. \tag{17}$$

To calculate B_p by means of (17) was tedious and therefore an approximation was made based on the use of β' the visual attenuation coefficient of the ozone layer, of β the visual attenuation coefficient of the atmosphere, and of I_0' and I_0 the illuminations of sunlight on the top of the ozone layer and the top of the atmosphere, respectively. One has

$$\beta'=\int_0^\infty i_{0\lambda}'\psi_\lambda\beta_\lambda'd\lambda\Big/\int_0^\infty i_{0\lambda}\psi_\lambda d\lambda, \tag{18}$$

$$\beta=\int_0^\infty i_{0\lambda}\psi_\lambda d\lambda\Big/\int_0^\infty i_{0\lambda}\psi_\lambda d\lambda, \tag{19}$$

$$I_0'=\int_0^\infty i_{0\lambda}'\psi_\lambda d\lambda, \tag{20}$$

$$I_0=\int_0^\infty i_{0\lambda}\psi_\lambda d\lambda. \tag{21}$$

From (13), (18), (20), and (21), one finds to a close approximation

$$I_0=I_0'\exp(-\beta'\sec\zeta). \tag{22}$$

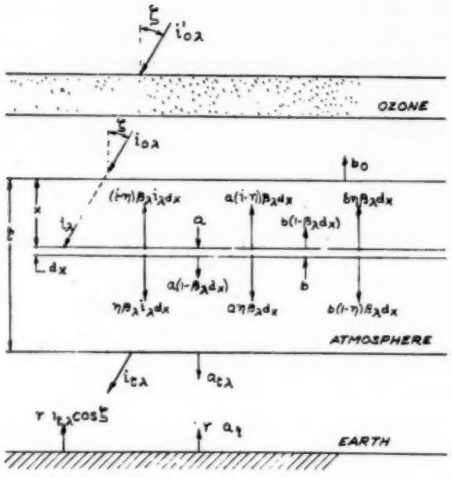

FIG. 12. Geometry of multiple scattering.

When (18)–(22) were introduced into (16) one obtained approximately

$$B_p=\frac{3}{4\pi}I_0'\exp(-\beta'\sec\zeta)\frac{1+\dfrac{1-\delta}{1+\delta}\cos^2\phi}{3+\dfrac{1-\delta}{1+\delta}}\times\frac{e^{-\beta x_1\sec\zeta}-e^{-\beta x_1\sec\xi}}{1-\sec\zeta\cos\xi},\quad(23)$$

which is the expression used to calculate B_p.

Analysis showed that (23) departed from the exact expression (17) by an amount that increased with ξ and ζ to about 4 percent for $\xi=\zeta=75°$. It was concluded that (23) was sufficiently accurate for the present purpose.

Calculation of B_m

In order to calculate B_m one must evaluate the flux of light, other than the direct rays of the sun, which impinges upon the atmosphere along the line of sight of the observer. The flux comes from the surrounding atmosphere and from the surface of the earth, which in turn owe their illumination to the sunlight. As the rays of the sun pass into the atmosphere a fraction of their light is scattered downward and upward, thereby giving rise to two streams of diffuse radiation, one stream of intensity a moving downward and stream of intensity b moving upward. The rays of the sun finally pass through the atmosphere and are assumed to be diffusely reflected upward by the surface of the earth. The reflected light passes through the atmosphere, a fraction of it is scattered and contributes to the a and b streams. The calculation of the values of a and b is given below. Actually, the calculation has already been made in an earlier paper[8] by dealing first with the case of a scattering and absorbing medium and then reducing the results to the simpler case of a scattering and non-absorbing medium. For the sake of clearness the direct calculation of this simpler case is given here.

In Fig. 12 is shown a vertical section of the atmosphere with, to avoid confusion, the ozone region separated from the "top" of the atmosphere, and the "bottom" of the atmosphere separated from the surface of the earth. Let t be the total thickness of the atmosphere and x the distance from the top of the atmosphere to a horizontal layer of thickness dx. The intensity of sunlight on the top of the atmosphere is $i_{0\lambda}\cos\zeta$ and on the layer is $i_\lambda\cos\zeta$. Assuming no true absorption on passage through the layer dx an amount $\beta_\lambda i_\lambda\cos\zeta dx/\cos\zeta=\beta_\lambda i_\lambda dx$ is scattered. A fraction η of this is scattered downward and contributes to the a stream and a fraction $(1-\eta)$ is scattered upward and contributes to the b stream. Similarly for the a and b streams in their passage through the layer. The details of the various attributes are shown in Fig. 12. In general a and b are functions of λ. For a Rayleigh atmosphere $\eta=\frac{1}{2}$ for both diffuse and collimated radiation.

Then

$$da=a(1-\beta_\lambda dx)+a\eta\beta_\lambda dx+b(1-\eta)\beta_\lambda dx+\eta\beta_\lambda i_\lambda dx-a,\quad(24)$$

$$db=b-b(1-\beta_\lambda dx)-b\eta\beta_\lambda dx-a(1-\eta)\beta_\lambda dx-(1-\eta)\beta_\lambda i_\lambda dx,\quad(25)$$

which reduce to

$$da/dx=(-a+b)(1-\eta)\beta_\lambda+\eta\beta_\lambda i_\lambda,\quad(26)$$

$$db/dx=(-a+b)(1-\eta)B_\lambda-(1-\eta)\beta_\lambda i_\lambda.\quad(27)$$

Let subscripts 0 and t denote the values of the quantities at the upper and lower surfaces of the

[8] E. O. Hulburt, J. Opt. Soc. Am. **33**, 42–45 (1943).

atmosphere, respectively. Let r be the diffuse reflectivity of the surface of the earth and assume that r is independent of λ. We note that

$$i_\lambda = i_{0\lambda} e^{-\beta_\lambda x \sec\zeta}, \quad i_{t\lambda} = i_{0\lambda} e^{-\beta_\lambda t \sec\zeta}, \tag{28}$$

and

$$a_0 = 0, \quad b_t = r(a_t + i_{t\lambda} \cos\zeta). \tag{29}$$

Introducing (28) and (29) into (26) and (27) and solving eads to

$$a = \frac{i_{0\lambda} \cos\zeta}{1+gt}\{(1+gt)(C-CX) - gx(C-CT+T)\}, \tag{30}$$

$$b = \frac{i_{0\lambda} \cos\zeta}{1+gt}\{(1+gt)(C-CX+X) - (gx+1-r)(C-CT+T)\}, \tag{31}$$

where

$$g = (1-r)(1-\eta)\beta_\lambda, \quad X = e^{-\beta_\lambda x \sec\zeta}, \quad C = \eta + (1-\eta)\cos\zeta, \quad T = e^{-\beta_\lambda t \sec\zeta}. \tag{32}$$

Equations (30) to (32) are the same as Eqs. (22) to (27) of the former paper.[8]

Referring to Fig. 11 an element $dx/\cos\xi$ of the atmosphere at a point P on the line of sight of the observer is subjected to downward and upward streams of radiation of intensities a and b, respectively. Assuming that these streams of radiation are unpolarized and are uniformly distributed over the upper and lower hemispheres, the amount scattered by the element in any direction per unit solid angle is

$$\frac{1}{4\pi}\beta_\lambda(a+b)\sec\xi dx. \tag{33}$$

The intensity which reaches the observer is (33) multiplied by

$$e^{-\beta_\lambda(x_1-x)\sec\xi},$$

and the total intensity $i_{m\lambda}$ along the line of sight is the integral from 0 to x_1. Then

$$i_{m\lambda} = \frac{1}{4\pi}\int_0^{x_1} \beta_\lambda(a+b)\sec\xi e^{-\beta_\lambda(x_1-x)\sec\xi} dx, \tag{34}$$

and the brightness B_m is

$$B_m = \int_0^\infty \psi_\lambda i_{m\lambda} d\lambda. \tag{35}$$

It was tedious to calculate B_m from (35) and a sufficiently close approximation was obtained by replacing β_λ and $i_{0\lambda}$ in (30)–(32) by β of (19) and I_0 of (21), and calculating B_m from

$$B_m = \frac{1}{4\pi}\int_0^{x_1} \beta(a+b)\sec\xi e^{-\beta(x_1-x)\sec\xi} dx. \tag{36}$$

The direct integration of (36) resulted in a cumbersome expression. A further approximation was made which led to little departure from exactness and to considerable simplification. Since in the present investigation β was small, average values of a and b were used and taken from under the integral sign. Let a_1, a_0, and b_1,

TABLE I. Theoretical sky brightness and polarization at 10,000 feet for clear air. $\beta = 0.017$.

Zenith angle of sun $\zeta = 0°$

ξ	B_p	B_m (Any Z)	γ
0°	133	24	1.00
10°	133	24	0.98
20°	133	25	0.90
30°	134	27	0.79
40°	137	31	0.66
50°	145	36	0.52
60°	163	46	0.39
70°	204	64	0.28
75°	246	81	0.24
80°	330	111	0.21

Zenith angle of sun $\zeta = 30°$

	Z=0°			Z=45°			Z=90°			Z=135°			Z=180°		
ξ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ
0°	116	21	0.79	116	21	0.79	116	21	0.79	116	21	0.79	116	21	0.79
10°	126	21	0.91	123	21	0.86	116	21	0.77	110	21	0.69	107	21	0.66
20°	138	22	0.98	132	22	0.90	117	22	0.72	105	22	0.57	101	22	0.51
30°	151	24	1.00	141	24	0.89	119	24	0.64	102	24	0.44	97	24	0.38
40°	166	27	0.98	152	27	0.83	123	27	0.54	103	27	0.33	97	27	0.27
50°	188	32	0.91	169	32	0.74	133	32	0.43	111	32	0.24	106	32	0.20
60°	221	40	0.79	196	40	0.62	154	40	0.33	132	40	0.18	130	40	0.17
70°	283	56	0.66	251	56	0.50	199	56	0.24	184	56	0.17	189	56	0.20
75°	338	71	0.59	300	71	0.44	242	71	0.22	233	71	0.18	246	71	0.23
80°	440	98	0.52	401	98	0.41	325	98	0.19	329	98	0.20	352	98	0.27

Zenith angle of sun $\zeta = 60°$

	Z=0°			Z=45°			Z=90°			Z=135°			Z=180°		
ξ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ
0°	79	14	0.35	79	14	0.35	79	14	0.35	79	14	0.35	79	14	0.35
10°	90	14	0.50	86	14	0.43	80	14	0.34	74	14	0.26	72	14	0.24
20°	105	15	0.64	97	15	0.54	82	15	0.33	72	15	0.19	70	15	0.16
30°	125	16	0.78	111	16	0.61	86	16	0.30	75	16	0.15	74	16	0.14
40°	150	18	0.90	129	18	0.66	94	18	0.26	83	18	0.14	85	18	0.16
50°	185	21	0.98	154	21	0.68	107	21	0.22	100	21	0.15	108	21	0.24
60°	233	26	1.00	189	26	0.65	128	26	0.19	130	26	0.20	149	26	0.35
70°	325	37	0.98	260	37	0.62	176	37	0.16	197	37	0.28	237	37	0.49
75°	404	46	0.94	321	46	0.58	220	46	0.15	260	46	0.32	317	46	0.57
80°	541	64	0.90	429	64	0.55	300	64	0.14	371	64	0.37	460	64	0.64

Zenith angle of sun $\zeta = 75°$

	Z=0°			Z=45°			Z=90°			Z=135°			Z=180°		
ξ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ	B_p	B_m	γ
0°	60	8	0.16	60	8	0.16	60	8	0.16	60	8	0.16	60	8	0.16
10°	67	9	0.26	65	9	0.22	61	9	0.15	58	9	0.11	58	9	0.10
20°	78	9	0.40	72	9	0.30	63	9	0.15	60	9	0.10	60	9	0.10
30°	94	10	0.55	84	10	0.38	68	10	0.14	65	10	0.11	69	10	0.16
40°	117	11	0.71	100	11	0.47	75	11	0.13	77	11	0.16	84	11	0.25
50°	150	13	0.84	123	13	0.53	87	13	0.12	95	13	0.21	111	13	0.39
60°	200	16	0.94	159	16	0.60	109	16	0.11	129	16	0.29	157	16	0.54
70°	284	22	0.99	221	22	0.59	149	22	0.10	190	22	0.37	240	22	0.70
75°	366	28	1.00	283	28	0.58	198	28	0.10	252	28	0.42	322	28	0.77
80°	500	39	0.99	386	39	0.57	262	39	0.10	357	39	0.46	459	39	0.84

TABLE II. Thickness of atmosphere x_1.

Altitude above surface, feet	x_1 km	Altitude above surface, feet	x_1 km
0	8.00	30000	2.37
2000	7.45	32000	2.16
4000	6.91	34000	1.97
6000	6.40	36000	1.79
8000	5.94	38000	1.63
10000	5.50	40000	1.48
12000	5.08	42000	1.35
14000	4.70	44000	1.23
16000	4.34	46000	1.11
18000	4.00	48000	1.00
20000	3.68	50000	0.91
22000	3.38	52000	0.83
24000	3.10	54000	0.75
26000	2.84	56000	0.68
28000	2.60	60000	0.62

b_0 be the values of a and b at x_1 and 0, respectively, from (29) $a_0=0$. The average values over the interval of integration are $a_1/2$ and $(b_1+b_0)/2$. Then (36) reduces to

$$B_m=\frac{1}{8\pi}(a_1+b_1+b_0)(1-e^{-\beta x_1 \sec\xi}), \qquad (37)$$

which is the expression used to calculate B_m.

The assumption made in obtaining (33), namely, that the radiation streams a and b were unpolarized and uniformly distributed over their respective hemispheres deserves comment. Although the streams are in general not completely unpolarized and not uniformly distributed, consideration indicates that the assumption introduces an error in B_m of probably less than 30 percent. Since B_m is usually less than $\frac{1}{4}$ of B (see Table I) it will produce errors in B probably less than 10 percent. However, the assumption has the effect of obliterating the points of "neutral," i.e., zero, polarization which are observed in the sky. In other words, the present theory does not lead to the neutral points, and the cause of this inadequacy appears to be in the above assumption. This is of negligible importance in the present connection, since the neutral points are in the region within 30° of the sun, a region where the sky is but slightly polarized anyway.

NUMERICAL VALUES

The following numerical values were used in the theoretical equations.

1. Atmospheric thickness x_1.—The equivalent vertical thickness x_1 of the atmosphere above each altitude above the surface, reduced to km of air at N.T.P., was calculated from the pressure altitude values[9] of the "U.S. Standard Atmosphere." The values of x_1 for various altitudes above the surface are given in Table II.

2. Earth reflectivity r.—The diffuse reflectivity, or albedo, r of the surface of the earth was measured in flight at low altitude by comparing the brightness of the surface with that of a horizontal white plaque exposed to the sun and sky. The following values of r were obtained:

forest	4 to 10 percent
green fields	10 to 15
dry grass fields	15 to 25
dry ploughed fields	20 to 25
bay and rivers	6 to 10
sea	3 to 7

Snow, clouds, and low altitude haze and fog below the observer increase r; gloss of vegetation covered fields and shine of water modify r. The flights of the present program occurred in moderately clear winter weather, no snow, and mostly over land. r was observed to be between 0.10 and 0.25 during all the flights. We use

$$r=0.20.$$

3. Polarization defect δ *of air.*—For pure air the observed values of δ were*: Rayleigh, 0.042, 0.050; Cabannes, 0.041; Raman, 0.0437; Rao, 0.0415. We use

$$\delta=0.04.$$

4. Absorption coefficient β' *of atmospheric ozone.*—The values of β_λ' for atmospheric ozone in the Chappuis band of the visible spectrum have been given by Fowle[10] for the atmosphere above Mount Wilson, the total amount of ozone in a vertical column being 0.32 cm at N.T.P. From the values β_λ', together with standard values[11] of ψ_λ and Abbot's[12] values of $i_{0\lambda}'$, the coefficient β' was calculated by means of (18). It was

[9] W. G. Brombacher, National Advisory Committee for Aeronautics, Report No. 538 (1935).

* Reference 4, page 384.

[10] F. E. Fowle, Smithsonian Miscellaneous Collections **81**, 1 to 27 (1929).

[11] *Smithsonian Physical Tables* (1932), eighth edition.

[12] C. G. Abbot, *Measurement of Radiant Energy* (McGraw-Hill Book Company, Inc., New York, 1937); see also A. C. Hardy, *Handbook of Colorimetry* (Massachusetts Institute of Technology Press, Cambridge, 1936), p. 18.

$\beta' = 0.023$. Fowle noted variations amounting to about 0.01 in β' caused by changes in the amount of ozone. Changes of this amount will cause changes of less than 3 percent in the present calculations. The amount of atmospheric ozone during the present experiments was not known. We take for a vertical column through the atmosphere

$$\beta' = 0.023.$$

5. Attenuation coefficient β of air.—β_λ was calculated throughout the visible spectrum from (11) with the values of $i_{0\lambda}$ determined from Abbot's values[12] of $i_{0\lambda}'$ by means of (13) and with standard values[11] of μ_λ. β was then calculated from (19) and was 0.101 per atmosphere or 0.0126 per km of air at N.T.P. The correctness of this theoretical value of β for pure air has never been tested rigorously. However, observations of the atmosphere above Mount Wilson corrected for haze and dust were shown[13] to agree with the Rayleigh formula within a few percent for wave-lengths from about 4500 to 3600A outside of the ozone bands.

We take for *pure* air the value

$$\beta = 0.0126 \text{ per km.}$$

It may be stated in advance that we have encountered an atmosphere that is not *pure* air, and have been concerned with a "clear" atmosphere, defined in an arbitrary way, for which

$$\beta = 0.017 \text{ per km.}$$

6. Illumination of sun I_0'.—The solar illumination I_0' outside of the atmosphere was determined in several ways.

(a) Calculation from (20) with Kimball's[14] values of the solar illumination at ground level for an average clear atmosphere at Washington, D. C., and with Abbot's[12] average spectral transmission coefficients of the Washington atmosphere, reduced to a zenith sun, gave $I_0' = 13600$ footcandles. The value is open to some uncertainty since it is not known whether the "average clear atmosphere" of Kimball was the same as that of Abbot. However, the value appears to be the most accurate available at present.

(b) Calculation from Abbot's[12] solar spectral energy curve and solar constant gave $I_0' = 11300$ footcandles. It was apparent, however, that the value was open to considerable uncertainty because of the lack of spectral energy data in the ultraviolet and infra-red regions of the spectrum.

(c) During the stratosphere flight[15] of Explorer II a value $I_0 = 12000$ footcandles was obtained when the zenith angle of the sun was about 60°. The thickness of the ozone was observed to be 1.9 mm from measurements made during the flight.[16] These values in (22) gave $I_0' = 12300$. The value seems low.

It is possible that I_0' may not be constant since long and short period fluctuations of a few percent in various regions of the visible solar spectrum have been mentioned.[17] We take

$$I_0' = 13{,}600 \text{ footcandles.}$$

The value refers to the mean solar distance.

COMPARISON OF EXPERIMENT AND THEORY

Since haze was always present at the surface there was little point in comparing the theory of a pure air atmosphere with surface observations. Therefore, the theoretical calculations were made for an altitude of 10,000 feet on the idea that at this altitude the atmosphere above the observer might approach pure air, as indeed it was found to do although not completely.

The numerical values for pure air, $\delta = 0.04$, $\beta = 0.0126$, $\beta' = 0.023$, $\tau = 0.20$, and $I_0' = 13{,}600$, were introduced into (5), (23), and (37), and B_p, B_m, and γ were calculated for an altitude of 10,000 feet, the values being listed in a table which is not printed here. The values were plotted in several families of curves, and from the curves, which are not included here, the pure air theoretical values of B, $B_\perp$, $B_{||}$, and γ could

[12] F. E. Fowle, Astrophys. J. **38**, 392–406 (1913); **40**, 435–442 (1914).

[14] H. H. Kimball, Trans. Ill. Eng. Soc. **16**, 255–283 (1921).

[15] R. P. Teele, Nat. Geog. Soc. U. S. Army Air Corps Stratosphere Flight of 1935, pages 133–138.

[16] B. O'Brien, F. L. Mohler, and H. S. Stewart, Nat. Geog. Soc. U. S. Army Air Corps Stratosphere Flight of 1935, pages 70–93.

[17] C. G. Abbot, F. E. Fowle, and L. B. Aldrich, Annals Astrophys. Obs. Smithsonian Institution **3**, 135 (1913); **4**, 17 and 207 (1922); C. G. Abbot, Smithsonian Miscellaneous Collections **101**, No. 5, preface (1941).

TABLE III. Zenith brightness B at 10,000 feet.

Flight	Observed B	Theoretical pure air B	Percent difference	β
1	153 c ft.$^{-2}$	96 c ft.$^{-2}$	59	0.0201
2	130	89	46	.0184
3	103	73	41	.0178
4	124	87	42	.0179
5	(101)	87	(16)	(.0146)
6	114	69	55	.0195
7	87	63	38	.0174
8	91	68	34	.0169
9	83	64	30	.0164
11	92	69	33	.0168
12	110	73	51	.0190

be read off for any point in the sky for solar altitudes between 15° and 90°. The pure air values of B obtained in this way were found in all cases to be less than the observed values at 10,000 feet by amounts ranging from 30 to 60 percent. This indicated that the atmosphere above 10,000 feet was not pure air but contained scattering particles in addition to the air molecules.

Thereupon, for reasons given in the following paragraphs, the attenuation β was increased to 0.017, an increase of about 35 percent over the pure air value 0.0126, and B_m, B_p, B, and γ were recalculated using the numerical values $\delta=0.04$, $\beta=0.017$, $\beta'=0.023$, $r=0.20$ and $I_0'=13{,}600$. The new values of B_m, B_p, and γ are given in Table I; B_m and B_p exceeded the pure air values by 30 to 37 percent, the exact amount depending on ζ and ξ. The values of γ of Table I were approximately the same as the pure air values. With $\beta=0.017$ new theoretical curves were calculated for comparison with the observations and are plotted in the dotted curves of Figs. 1 to 10. An atmosphere above 10,000 feet for which $\beta=0.017$ is defined as "clear."

Referring to the sky data at 10,000 feet of Figs. 4, 5, 9, and 10 it is seen that the "clear" air theoretical values of B are not far from, and are usually less than, the observed values, and that the theoretical values of γ agree with the observed values, the deviations appearing to be accidental, due to conditions of flight.

Turning to the zenith values of B, $B_\perp$, $B_{\shortparallel}$, and γ at various altitudes of Figs. 2, 3, 7, and 8, the observed values were considerably above the "clear" air theoretical curves at low altitudes due to prevalent haze near the surface. It is seen that the effect of the haze was to add to the polarized zenith light a considerable amount of unpolarized light. This increased $B_{\shortparallel}$ and $B_\perp$ by approximately equal increments and caused corresponding increases in γ. With increasing altitude, particularly after the haze level was passed, the observed quantities approached the theoretical curves showing that the observer was rising above the haze into purer air. In five of the flights, of which Figs. 2 and 3 are an example, they were approximately coincident with the theoretical curves at, or below, 10,000 feet showing that the observer had reached "clear" air for which β was 0.017. In seven of the flights, of which Figs. 7 and 8 are an example, an aureole was visible around the sun and the observed values were greater than the theoretical values showing that the air above 10,000 feet on those occasions contained large particle haze.

In none of the flights did the observer reach theoretically "pure" air, i.e., air for which $\beta=0.0126$. This is brought out in Table III which gives the observed and pure air theoretical values of B at the zenith and the percent difference for an altitude of 10,000 feet. The five lowest values of the difference averaged 0.30, corresponding to an increase of β to 0.017. The low value of B, and hence the small percent difference of Flight 5, was probably due to observational error, for the other data of this flight did not indicate an unusually low sky brightness. In the last column of Table III are given the values of β required to obtain agreement between theory and observation.

The only flight much above 10,000 feet was Flight 9 of the present experiments. In this case clouds interfered with an extended program of measurement, but a few values of B for the zenith sky were obtained at 18,800 feet. These were about 30 percent above the theoretical values for pure air. This meant that β for this case was about 0.017, and hence that the optical purity, or rather impurity, of the air above 18,800 feet was the same as that of the air above 10,000 feet for the condition described as "clear." It would be important to obtain more observations at altitudes above 10,000 feet.

Measurement of the solar illumination I are

plotted in Figs. 1 and 6. In the figures the dotted theoretical curves were calculated from the relation

$$I = I_0' \exp(-\beta' \sec\zeta) e^{-\beta x \sec\zeta}, \qquad (32)$$

where $I_0' = 13{,}600$, $\beta' = 0.023$, and $\beta = 0.017$, which are the values used for calculating the sky brightness and polarization for clear air. It is seen that the theoretical curve approaches agreement with the observations at the higher altitudes for the "clear" air case of Fig. 1, and deviates from the observations for the "less clear" air case of Fig. 7 as expected. At the lower altitudes the observed values of I were below the theoretical curves due to increasing haze near the surface.

It is of interest to compare the observations of Abbot, Fowle, and Aldrich[17] with the present results. They measured the transmission for various wave-lengths in the visible spectrum through a clear atmosphere above Mount Wilson (altiude 5680 feet). Calculation from their average data of 1905 and 1906 gave $\beta = 0.016$, in good agreement with the present result 0.017. The agreement is considered to indicate that the atmosphere above Mount Wilson described as "clear" was of about the same clarity as the atmosphere above 10,000 feet of the present investigation also described as "clear."

The conclusion that the attenuation and scattering of the "clear" air above 10,000 feet was about 35 percent above that of theoretically pure air, is not appreciably disturbed by uncertainties in the value of the earth reflectivity $r = 0.20$ used in the theoretical calculations. This may be seen from Table IV which gives the theoretical clear air ($\beta = 0.017$) values of B_p, B_m, B, and γ for r from 0 to 1 for two points in the sky. r was observed to be between 0.10 and 0.25 during the present flights, and Table IV shows that for this range of r the variances in B and γ were less than about 10 percent, respectively, from their values for $r = 0.20$. Furthermore, the value of γ of Flight 11 at the point of maximum polarization, taken purposely for this comparison, was 0.13, which agreed well with the value 0.14 of Table IV for $r = 0.20$. Table IV suggests, however, that B and γ might be noticeably increased by large areas of white clouds or fresh snow beneath the observer, for in these cases r might be greater than 0.4. Over the sea with a clear atmosphere r is usually less than 0.10. This would result in slightly lowered values of B and γ.

TABLE IV. Sky values at 10,000 feet as a function of r for "clear" air, $\beta = 0.017$.

	$\zeta = 60°$		$\xi = 30°$		$Z = 180°$	
r	0	0.2	0.4	0.6	0.8	1.0
B_p	74	74	74	74	74	74
B_m	5	16	25	36	47	56
B	79	90	99	110	121	130
γ	0.08	0.14	0.20	0.24	0.29	0.34
	$\zeta = 60°$		$\xi = 0°$			
r	0	0.2	0.4	0.6	0.8	1.0
B_p	79	79	79	79	79	79
B_m	4	14	22	32	42	50
B	83	93	101	111	121	129
γ	0.31	0.35	0.39	0.42	0.45	0.47

The result is that we assume that the "clear" atmosphere above 10,000 feet is an atmosphere composed of pure air molecules with a number of foreign particles of size small with respect to λ sufficient to increase the light scattering and attenuation of pure air by 35 percent. This amounts to a definition of a "*clear atmosphere above 10,000 feet.*" True, the definition is arbitrary but, being based on experiment, is practical and permits one to say that if the additional light scattering is greater or less than 35 percent of that of pure air, the atmosphere is less clear, or clearer, than "*clear*," respectively.

Now an atmosphere whose attenuation and scattering is only about 35 percent greater than that calculated for pure air molecules is very pure indeed in comparison with the amount of optical contamination of the usual clear atmosphere at sea level. It is possible that the theory may be incorrect to a certain extent and that a portion of the 35 percent difference may be unreal. However, for the present we regard the entire 35 percent difference as real and due to scattering particles. For want of a better term we may refer to these as "haze" particles, in spite of the fact that because of their extreme tenuousness the appellation "haze" may be misleading, since haze ordinarily refers to an easily perceptible obscurity in the lower atmosphere.

The data, of which Figs. 4 and 9 are a sample, showed that the values of B at 10,000 feet, especially toward the sun, were in many cases greater than the theoretical curves for $\beta = 0.017$.

This was taken to mean that in these cases there was additional haze which varied from flight to flight. A detailed analysis of the differences between the observed and theoretical values indicated that the haze particles were not all of the same size and that the size distribution varied from flight to flight. This seemed reasonable from a meteorological standpoint, for one would expect the composition of the haze, as water droplets, dust, or whatever it was, to vary with the previous history and condition of the atmosphere. Small haze particles will produce no visible aureole, and large particles may produce a visible aureole. Apparently both types of particles were encountered in the flights of the present experiments.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 2 FEBRUARY, 1947

A Hollow Cathode Source Applicable to Spectrographic Analysis for the Halogens and Gases*

J. RAND MCNALLY, JR., GEORGE R. HARRISON, AND EUGENE ROWE**
Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 1, 1946)

Investigations aimed at developing methods of spectrographic quantitative analysis of the halogens and non-metals have been carried out, with emphasis on analysis for fluorine. The relative sensitivities of fluorine spectrum lines in the vacuum ultraviolet region and in the visible region were studied, and it was found that certain visible lines are, at present, more sensitive than the theoretical *raies ultimes* of fluorine. The greatest absolute sensitivity obtained, using a specially designed hollow cathode source, involved detection of about 0.01 microgram of fluorine, while the greatest concentrational sensitivity was slightly less than one part per million. Chlorine and sulphur were readily detected in amounts as small as 0.20 microgram and 1.0 microgram, respectively, in samples weighing 20 milligrams. Sample handling time was reduced sufficiently to permit routine analyses for these elements.

INTRODUCTION

THE purpose of this investigation was to develop a method of making quantitative determinations of fluorine and similar elements of high excitation potential when present in samples in extremely low concentrations, of the order of one part per million. Although it is commonly stated in the literature that spectrographic determinations of halogens and gases are not feasible with the usual procedures of emission spectroscopy, it was recognized that the spectrographic method, properly applied, should permit both high sensitivity and routine methods of analysis for these elements.[1]

Some success in analyzing for fluorine has been achieved in the past using the spectrographic method. Utilizing the high sensitivity of the band head of CaF_2 at 5291.0A, Papish, Hoag, and Snee[2] obtained an absolute sensitivity of ten micrograms of fluorine. Churchill[3] detected fluorine in concentrations as low as ten parts per million although he admitted the possibility of large errors.

In 1929 de la Roche[4] reported results on the use of sparks in atmospheres containing low concentrations of non-metallic compounds. The highest sensitivity he obtained for fluorine was 100 p.p.m. of SiF_4 in oxygen. Moritz[5] discussed the difficulties of analyzing for the elements H, O, N, F, Cl, Br, I, S, Se in this same year. In 1932 Gatterer[6] studied mixtures of gases and discussed the possibility of practical gas analyses with the spectrograph. Klauer[7] in 1934 described apparatus and methods for determining H_2 in argon or helium and argon or helium in molecular hydrogen. Konovalov and Frisch[8] used the hollow cathode to detect nitrogen in argon to a few tenths of a percent. Heyes[9] published a report in 1935 on the quantitative spectrographic analysis of mixtures of hydrogen, nitrogen, oxygen, and helium in which he considered the last lines and the detectable limits. Paul[10] gave a limit of detection of 0.1 mg for the analysis of fluorine as based on the intensities of silicon lines, heating the sample with SiO_2 and H_2SO_4 to form SiF_4. Schultes[11] considered apparatus and methods for determining the spectrographic analysis of gases.

The lack of sensitivity of the halogens and gases in ordinary laboratory sources arises from two major causes: first, the presence in the sample of other elements having low ionization

* This document is based on work performed under Contract No. W-7405-Eng-175 for the Manhattan Project at the Massachusetts Institute of Technology.

** Now with the B. F. Goodrich Company, Akron, Ohio.

[1] For a general discussion of this problem consult G. R. Harrison, "Practical possibilities in spectrographic analysis," Metals and Alloys (November, 1936).

[2] J. Papish, L. E. Hoag, and W. E. Snee, Ind. Eng. Chem. Anal. Ed. **2**, 263 (1930).

[3] H. V. Churchill, Ind. Eng. Chem. **23**, 996 (1931).

[4] B. de la Roche, Bull. Soc. Chim. **45**, 922 (1929).

[5] H. Moritz, Metall. Erz. **26**, 247 (1929).

[6] A. Gatterer, Physik. Zeits. **33**, 64 (1932).

[7] F. Klauer, Ann. d. Physik **20**, 145 (1934).

[8] V. A. Konovalov and S. E. Frisch, J. Tech. Phys. (USSR) **4**, 523 (1934).

[9] J. Heyes, Zeits. f. physik. Chemie **A172**, 95 (1935).

[10] W. Paul, Zeits. f. angew. Chemie **49**, 901 (1936).

[11] W. Schultes, Zeits. ver. deut. Ing. **81**, 373 (1937).

TABLE I. Strongest lines of certain first spectra.

Atomic number	Element	Principal line*	E.P. (volts)	I.P. (volts)
6	C	1656.994	7.45	11.22
7	N	1134.973	10.87	14.48
8	O	1302.183	9.48	13.55
9	F	954.817	12.92	18.6
15	P	1774.942	6.95	11.1
16	S	1807.37	6.83	10.3
17	Cl	1347.2	9.16	12.96
33	As	1890.53	6.53	—
34	Se	1960.87	6.29	9.5
35	Br	1488.40	8.29	11.80

* Wave-length from M.I.T. card catalog.

potentials diminishes the excitation of elements having relatively high excitation potentials; second, the principal spectrum lines of most nonmetals lie in the wave-length region below 2000A, which requires vacuum apparatus and special techniques. These difficulties are implied in the statement by Meggers,[12] "The notable success of practical spectrographic analysis for seventy elements is related to the fact that these elements have ionization potentials less than ten volts, and strongest lines greater than 2000A in wave-length." This is illustrated by Table I, which gives the principal lines of several halogens, gases, and metalloids, as well as the excitation potentials for these lines and the ionization potentials for the elements. By way of comparison, the average ionization potential for 48 other elements, excluding the inert gases and hydrogen, is only 6.99 volts, with an average excitation potential for the principal lines of only 3.78 volts.[12]

The problem here attacked involved two distinct phases: first, the selection of the most sensitive spectrum lines from a practical spectrographic viewpoint; second, the selection of source conditions which would permit a high average excitation of the atoms or ions of interest, with a possible reduction in the preferential excitation of other atoms or ions.

SENSITIVE LINES

Although the spectra of the halogens and gases are relatively simple, their wave-length range, even for the neutral atoms, extends from well in the infra-red to the extreme vacuum ultraviolet. The *raies ultimes* or principal lines of the neutral atoms are found at wave-lengths much shorter than 2000A, as shown in Table I. The principal spectrum lines of fluorine in various stages of ionization are given in Table II, which lists the ion state, the wave-length[13] of the spectrum line in angstrom units, the relative line intensities as given by Edlen[13] and by Bowen[14] in high vacuum spark sources, and lastly, the electronic transition. All of these sensitive lines of fluorine fall well below the transmission limits of known solids and hence no prisms, lenses, or windows can be used in observing the lines.

The intensities of fluorine lines as given in Table II indicate that the F I lines are probably not very sensitive in vacuum sources. The type of electronic transition ($s \rightarrow p$) is also not as productive of true resonance lines as is the reverse case ($p \rightarrow s$). A somewhat similar situation is found in the electronic transitions for the resonance lines of oxygen, sulphur, chlorine, and bromine. It seemed quite probable, therefore, that in high vacuum sources one of the higher stages of ionization for these atoms would produce more sensitive lines than does the neutral atom. In order to obtain maximum sensitivity, however, the available excitation energy should be concentrated into as few ionization stages as possible, preferably the lower stages to reduce the "gas clean-up" of ions at the walls of the source.

The excitation of lines in the ordinary spectrographic region should be such as to favor the production of the spectrum of the neutral atom. In addition, transitions of the type $p \rightarrow s$ would be expected to produce quasi-resonance lines, particularly if they involve metastable states. The expected sensitive lines of this type for several similar atoms are shown in Table III, which gives the atom with its probably most

TABLE II. Principal lines of fluorine.

Atom	Wave-length (A)	Intensity (E)	Intensity (B)	Electronic transition
F I	954.817	1	7	$s^2p^5 - s^2p^4s$
F II	606.805	8	9	$s^2p^4 - sp^5$
F III	658.337	12	8	$s^2p^3 - sp^4$
F IV	679.217	16	5	$s^2p^2 - sp^3$

[12] W. F. Meggers, J. Opt. Soc. Am. **31**, 26 (1941).

[13] B. Edlen, Zeits. f. Physik **92**, 21 (1934); **94**, 48 (1935); **98**, 450 (1935).

[14] I. S. Bowen, Phys. Rev. **45**, 82 (1934).

sensitive line[15] in angstroms, its spark intensity[15] and the electronic transition. Since all of the cases listed are similar in their electronic interchange it should be expected that these red lines are of comparable intensity. The low probability of intersystem combinations for these atoms makes the lower energy state essentially a metastable level,[16] and hence, fairly high sensitivity with these lines should result providing photographic difficulties are not too great.

TABLE III. Quasi-resonance lines of certain first spectra.

Atomic number	Element	Wave-length (A)	Spark intensity	Electronic transition
8	O	7771.928	1000	$2p^33s-2p^33p$
9	F	6856.02	1000	$2p^43s-2p^43p$
16	S	9212.91	200	$3p^44s-3p^44p$
17	Cl	8375.97	40	$3p^44s-3p^44p$
35	Br	8272.46	70	$4p^45s-4p^45p$

EXCITATION METHODS

The methods of atomic excitation are fundamentally: (1) thermal excitation, (2) resonance absorption, (3) electronic or atomic bombardment (collisions of the first kind), and (4) atomic transfer of excitation energy (collisions of the second kind). Methods (1) and (2) are generally not adaptable to the production of emission spectra corresponding to high excitation states. The last two methods, however, are extremely important in the production of emission spectra of high excitation.

Electrons or ions accelerated through a potential field in a gas acquire kinetic energy of an amount which depends on the field intensity, the gas density and the collision cross section of the atoms composing the gas. Those electrons or ions which have sufficient translational energy on colliding with an atom may transfer all or part of this energy to excitation energy of the atom. Successive collisions of this type involving low energy electrons can lead to cumulative processes which may result in extremely high atomic excitation energies. In high voltage sparks between solid electrodes the initial breakdown with high speed electrons is usually followed by an arc or glow discharge which favors much lower average electron speeds. This initial phase is frequently enhanced by storing energy in condensers connected across the spark gap, thus producing a discharge of greater energy content. The "hardness" of the spark can also be increased by reducing the gas pressure to below 10^{-4} mm of Hg, each electron on the average falling through a greater potential drop between collisions because of its increased free path.

The electrodeless discharge tube, excited by a high frequency field, is also productive of high average energy of electrons and ions with accompanying high atomic excitation. Here, also, pressure of gas, tube geometry, and temperature as well as the electrical conditions markedly control the degree of excitation.

In all processes involving electronic bombardment there is considerable atomic bombardment. Collisions of the first kind between atoms play only a minor role in the excitation of the atomic energy states. Collisions of the second kind, in which one atom possessing excess energy transfers this energy in whole or in part to a second atom, constitute the most important phase of atomic bombardment. Those elements which have metastable energy states, such as the inert gases, greatly influence the probability of such interchanges because of the longer mean life of the atom in the metastable state. Probability laws pertaining to collisions of the second kind have been recognized: (1) the yield is greater when a minimum of excitation energy must be converted to kinetic energy (i.e., it is essentially a resonance phenomenon),[17] and (2) the total spin of the collision partners must remain constant.[18]

The necessity for a high average excitation in the source demands that the choice of a suitable source must depend on the action of either or both of these collision processes. The high vacuum or hot-spark source was selected as meriting special study for use in the vacuum ultraviolet region, and the hollow cathode source was chosen for the ordinary photographic region.

[15] G. R. Harrison, *M. I. T. Wavelength Tables* (John Wiley and Sons, Inc., New York, 1939).
[16] R. F. Bacher and S. Goudsmit, *Atomic Energy States* (McGraw-Hill Book Company, Inc., New York, 1932).
[17] J. Franck, Naturwiss. **14**, 211 (1926); H. Beutler and B. Josephy, Zeits. f. Physik **53**, 747 (1929); H. Kallmann and F. London, Zeits. f. physik. Chemie Abt. **B2**, 207 (1929).
[18] E. Wigner, Göttinger Nachrichten 375 (1927); H. Beutler and W. Eisenschimmel, Zeits. f. physik. Chemie Abt. **B10**, 89 (1930).

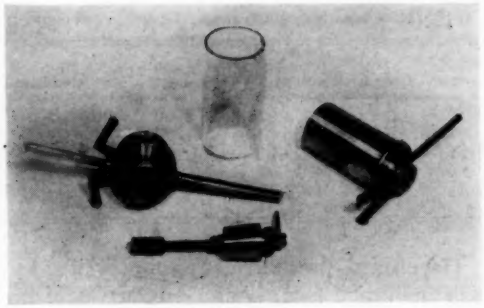

FIG. 1. The disassembled hollow cathode source (the glass tube is four inches long).

INVESTIGATIONS IN THE VACUUM ULTRAVIOLET

The excitation conditions of all high vacuum sources are such that the available excitation energy is distributed among several ion stages instead of being concentrated in a single stage (see Table II). In addition, low transmission of both solids and gases in the vacuum ultraviolet region is accompanied by a diminishing reflecting power of grating materials and a decreased sensitivity of photographic emulsions. All of these obstacles combine to lower the efficiency of the spectrographic method in this short wave-length region. The preliminary investigation was designed to test whether these difficulties were too great to permit the use of the theoretical *raies ultimes* for analysis purposes.

Most of the work in the vacuum ultraviolet region was carried out with a two meter grazing-incidence glass grating spectrograph, intense spectra being obtained at high dispersion from 200A to 2000A. Various photographic emulsions were tested for sensitivity and uniformity of response. Eastman 103–0 UV sensitive plates were finally selected as most practical. These commercially available plates are sensitive to at least 200A.

The "hot spark" source was operated at a high vacuum secured by an oil condensation pump connected through a large aperture stopcock to the main body of the source. This permitted a rapid evacuation of the gases evolved during sparking. Power was obtained from a 230 volt, 5 kw transformer capable of producing a peak voltage of about 70 kv. The sparking conditions were controlled by an external spark gap, a set of charging condensers of about 0.04 mf capacitance, and a current limiting resistor in the primary circuit. A mechanical interrupter, immersed in transformer oil, was also connected in the primary in order to control the hardness of the spark by permitting a more complete exhaustion of the evolved gases in the interval between sparks.

The sensitivity limit of the 830A lines of O II and O III using the hot spark source was found to be about 500 p.p.m. in metal oxides, with a probable ultimate limit near 300 p.p.m. Investigations on fluorine indicated a practical sensitivity in the neighborhood of 300 p.p.m.

Although special efforts were made to hold constant the experimental conditions on all exposures, the spectrum density varied quite markedly from exposure to exposure. In the hot spark source the electrodes are eaten away with noticeable rapidity, while the disintegration products clog the slit and form a conducting layer on the glass insulators of the high voltage electrodes. In addition, unit sample handling time amounts to several hours. This prohibitive exposure time, together with the relatively poor sensitivity and the non-reproducibility of conditions, emphasized the need for further developmental work before the extreme vacuum ultraviolet can be adapted to routine spectrographic analyses.

INVESTIGATIONS WITH THE HOLLOW CATHODE SOURCE

Schüler and Gollnow[19] suggested the application of the hollow cathode source to the analysis

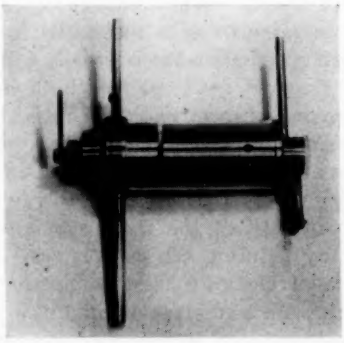

FIG. 2. The assembled hollow cathode source.

[19] H. Schüler and H. Gollnow, Zeits. f. Physik **93**, 611 (1935).

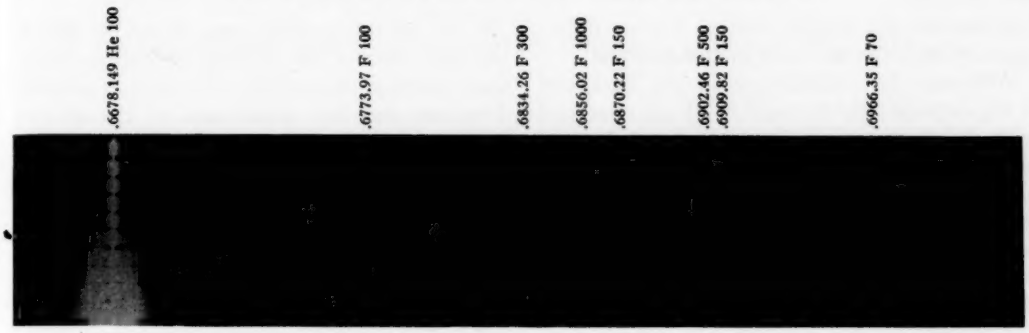

FIG. 3. Reproduction of a moving plate spectrogram using the hollow cathode source and a twenty milligram metal-oxide sample containing 1000 parts per million of fluorine.

of materials available in minute quantities. Our initial studies of several simple hollow cathode sources indicated the definite superiority of this method for the detection of the red lines of fluorine over the vacuum hot spark method for the extreme ultraviolet fluorine lines. A source similar in design to that recommended by Schüler was constructed and tested. Several modifications were found convenient to obtain more rapid analyses. The model illustrated in Figs. 1 and 2 is the product of numerous experiments with sources of varied designs.

The hollow cathode source contains a water-cooled anode, a stainless steel cathode connected to a ground joint and insulated electrically from the water-cooled cathode support, and a Pyrex glass tube securely waxed to both the cathode support and the anode. A glass window, waxed to the anode, permits light from the cathode to reach the spectrograph. To make possible viewing of the cathode discharge a small diagonal hole was drilled in the anode. With the exception of the cathode all metal parts are of yellow brass. The cathode, which was machined from a one-half inch diameter steel rod to eliminate all internal soldered joints, was drilled as far as the base of the cathode to accommodate an iron-nickel thermocouple. The insulator between cathode and cathode support provided sufficient heat conduction to prevent both over-heating of the cathode and air leaks through the ground joint. No stopcock grease was used on the ground joint but the outside was securely sealed with vacuum wax.

Sample handling time was reduced as much as possible to allow for the routine analysis of samples. The small source volume reduced the pumping time required following the introduction of each new sample. Helium, purified by continuous circulation through a liquid-air-cooled charcoal trap, was diverted through the evacuated source chamber, and the pressure was adjusted to an operating value of about one cm Hg. The whole procedure, together with the required exposure time of several minutes, involved at most about fifteen minutes per sample.

Power was obtained from a 2000-volt, 4 kva motor-generator set. The output voltage was controlled by a lamp bank in the field circuit, while current-limiting resistors totalling about 500 ohms were placed in series with the hollow cathode source. Suitable operating conditions were found to be about 800 volts across the tube at about 0.2 ampere.

A low dispersion Zeiss glass prism spectrograph was used for the preliminary investigations on fluorine. The limiting absolute sensitivity with the 6856A line was found to be about 0.01 microgram of fluorine. A 10-milligram sample containing 1 part per million of fluorine as sodium fluoride in a metal oxide matrix was used.

A strong background, due mainly to cathode incandescence, was noted. Background due to scattered light in the spectrograph and halo due to strong neighboring helium lines were eliminated by the use of Eastman I-F plates which have a fairly sharp cut-off at about 7100A, and also by the use of a filter which cut out all light of wave-length shorter than 6000A. With this

combination the intense helium line at 6678A appeared only weakly with long exposures.

Although the temperature at the back end of the cathode reached about 800°C, as measured with the thermocouple, the actual sample temperature was probably several hundred degrees higher, judging from its distinctly higher color temperature. This incandescence of the sample produced considerable non-uniform background when the sample was deposited on the back end of the cathode. By a proper deposition of the sample on the sides of the cathode this difficulty was greatly reduced.

The effect of the continuous background was reduced by using a thirty-five foot grating spectrograph in stigmatic (Wadsworth) mounting, with a dispersion of 3.4A/mm in the first order. Moving plate spectrograms as illustrated in Fig. 3 showed that a total burning time of several minutes was required to exhaust fluorine from the sample fully. A 20 mg sample containing 1000 p.p.m. of fluorine was used for Fig. 3. Six successive one minute exposures, two two-minute exposures, one five-minute exposure and one ten-minute exposure were made. No fluorine lines appeared after six minutes of passage of current at this particular concentration.

An ammonia-carbonate hypersensitizer[20] was next used on the Eastman I-F plates, and at wave-length 6856A a sensitivity gain of about 2.5 was obtained for low densities without any noticeable increase in background fog. With a hyper-sensitized plate, fluorine was detected in a 40-mg sample containing 1 p.p.m., whereas non-hypersensitized plates gave a sensitivity limit of 1 p.p.m. in a 100-mg sample.

Eight 20-mg metal oxide samples containing 30 p.p.m. fluorine as sodium fluoride were run under constant excitation conditions to check the reproducibility of the hollow cathode method. Although sample inhomogeneity probably accounts for some variation, the average deviation for the line intensities was found to be 15 percent, whereas the average deviation of the line intensity ratio F/He was only 6 percent. This indicates the dependence of the fluorine excitation on that of the helium. However, it should be noted that the fluorine excitation occurs at the expense of the helium excitation through collisions of the second kind, and hence such ratios are not absolute criteria of the true number of fluorine atoms present.

Investigation of the sensitivity of oxygen with hollow cathode excitation of 7771.928A showed that contamination of materials, insufficient purity of gas and other factors limited the sensitivity to about 100 p.p.m. Standard samples containing lead dioxide in freshly prepared carbon powder were used. Any adsorbed oxygen in the carbon powder proved to be less worrisome than other sources of contamination.

The presence of chlorine and sulphur in the fluorine standards permitted the simultaneous study of the limiting sensitivities of these elements. Their spectra are somewhat similar to fluorine, with the principal spectrum lines occurring in the vacuum ultraviolet although their quasi-sensitive lines occur somewhat further in the infra-red than do those of fluorine. However, one microgram of sulphur and 0.2 microgram of chlorine were detected in 20-mg samples. This corresponds to 50 p.p.m. and 10 p.p.m., respectively. No special emphasis was placed on the development of methods productive of high sensitivity for these elements, but it is believed that these figures could be substantially reduced.

Note added to proof: The halogens in higher concentration have been investigated by Johnson and Norman[21] with an a.c. condensed spark source. The limits of detection which they reached are 5000 p.p.m. for chlorine and 3000 p.p.m. for bromine, in a carbon dioxide atmosphere; and 1000 p.p.m. for fluorine and 700 p.p.m. for iodine, in air.

[20] D. MacDonald, Thesis for the Doctorate (Boston University, June, 1944).

[21] W. W. A. Johnson and D. P. Norman, Ind. Eng. Chem. Anal. Ed. **15**, 119 (1943).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 2 FEBRUARY, 1947

A Generalized Intensity Formula for a System of Retardation Plates

HSIEN-YÜ HSÜ,* M. RICHARTZ, AND YÜNG-K'ANG LIANG
Catholic (Fu-Jen) University, Peiping, China
(Received May 27, 1946)

INTRODUCTION

CONSIDER an optical system which consists of a polarizer, n retardation plates, and an analyzer, and arrange the elements of the system as shown in Fig. 1. We shall assume that monochromatic light is incident normally on each of the optical elements: namely, its direction of propagation (z axis), is perpendicular to the plane of the polarizer, the plates and the analyzer. We relate the "fast" direction of vibration p_i within the ith plate to the direction of vibration p (x axis) of the polarizer. Now, with respect to the x axis; suppose the orientation or azimuth of the ith plate p_i be ρ_i, with its phase difference $\delta_i \equiv 2\gamma_i$, and that of the analyzer be ϕ (Fig. 2). Let us denote the intensities of the polarized light emerging from the polarizer and the analyzer by I_0 and I_n, respectively.

For arbitrary values of the orientations and retardations the intensity formula for *two* plates was first given by Fresnel.[1] The general case of n plates was discussed by Tuckerman.[2] The intensity formula for *three* plates has been given in different forms for the sake of convenient computations.[3] Although the ideas involved in the derivation of the formula for *three* plates are quite simple and straightforward, the processes used to obtain the final result are long and cumbersome. Furthermore, a similar formula for *four* plates could hardely be anticipated from that for three plates in any of the given forms, and its direct derivation would involve too much care and work.

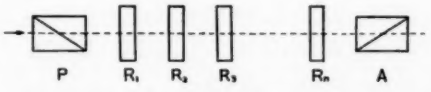

FIG. 1. Arrangement of the optical system: P, polarizer; A, analyzer; R_1, R_2, $\cdots R_n$, retardation plates.

Recently, a new calculus has been proposed by R. Clark Jones[4] in order to study some "equivalence" properties of the above optical system as shown in Fig. 1. From his consideration each of the retardation plates can be represented mathematically by the product of three two-by-two complex matrices.

In this paper we shall apply the new calculus to obtain the intensity formula for n plates. We shall first consider the retardation plates. Since each plate can be represented by a two-by-two complex matrix $\mathbf{M}_i$ n such plates will be equivalent to the product of n matrices, which is again a two-by-two complex matrix $\mathbf{M}^{(n)}$. We shall separate the complex matrix $\mathbf{M}_i$ into its real and imaginary parts. Their product matrix $\mathbf{M}^{(n)}$ can also be separated. The explicit, analytical expression for the separation of $\mathbf{M}^{(n)}$ shall be searched for, and proved by mathematical induction. Then, from the definition of the intensity of the light, we can establish the generalized intensity formula for a system of n retardation plates. In particular, when $n=3$, the formula is now presented in a "summed" form, which

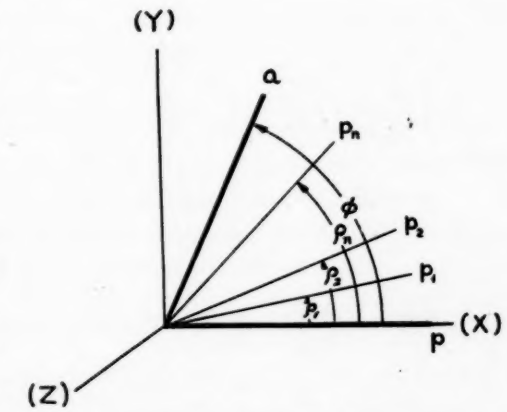

FIG. 2. Directions of vibration and their azimuths.

* Department of Mathematics, Yenching University, Peiping.

[1] A. Fresnel, Ann. Chem. Phys. [2], 17, 172 (1821).

[2] L. B. Tuckerman, Univ. Studies, Univ. Nebraska 9, 157 (1909).

[3] G. Szivessy and W. Herzog, Zeits. f. Instrumentenk. 57, 53 (1937); F. Gabler and P. Sokob, Zeits. f. Instrumentenk. 58, 302 (1938); R. D. Mindlin, J. Opt. Soc. Am. 27, 289 (1937).

[4] R. Clark Jones, J. Opt. Soc. Am. 31, 488 (1941).

appears at first sight quite different from those classical results as recorded in the literature. However, they can be shown to be equal to one another. Furthermore, the "summed" formula for *three* plates, when expanded, may be easily used to anticipate the intensity formula for *four* or more plates.

NOTATIONS IN THE NEW CALCULUS

The states of polarization may be completely defined by the x and y components of the electric vector of the light wave. Let the components of the electric vector be E_x and E_y. Jones has shown that the electric vector after passing through a retardation plate may be put mathematically as follows:

$$\mathcal{E}\equiv\begin{pmatrix}E_x\\E_y\end{pmatrix}=\begin{pmatrix}\cos\rho & -\sin\rho\\ \sin\rho & \cos\rho\end{pmatrix}\begin{pmatrix}e^{i\gamma} & 0\\ 0 & e^{-i\gamma}\end{pmatrix}\begin{pmatrix}\cos\rho & \sin\rho\\ -\sin\rho & \cos\rho\end{pmatrix}\begin{pmatrix}E_{x0}\\E_{y0}\end{pmatrix}. \quad (1)$$

Or we may simply write

$$\mathcal{E}=\mathbf{S}(\rho)\mathbf{G}(\gamma)\mathbf{S}(-\rho)\mathcal{E}_0 \quad (2)$$

where the product of the three two-by-two matrices

$$\mathbf{M}\equiv\mathbf{S}(\rho)\mathbf{G}(\gamma)\mathbf{S}(-\rho) \quad (3)$$

acts as the transformation of the electric vector. If there are n such plates, we shall have

$$\mathcal{E}_n=\mathbf{M}^{(n)}\mathcal{E}_0 \quad (4)$$

where

$$\mathbf{M}^{(n)}\equiv\mathbf{S}(\rho_n)\mathbf{G}(\gamma_n)\mathbf{S}(-\rho_n)\cdots\mathbf{S}(\rho_1)\mathbf{G}(\gamma_1)\mathbf{S}(-\rho_1). \quad (5)$$

NEW NOTATIONS

We shall study the transformation $\mathbf{M}$. For the sake of convenience it is desirable to separate the real and imaginary parts in $\mathbf{M}$. We denote

$$\mathbf{A}\equiv\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \quad \text{and} \quad \mathbf{B}\equiv\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

It is easy to show that $\mathbf{B}^2=\mathbf{A}$. Then we have

$$\mathbf{G}(\gamma)=\cos\gamma\,\mathbf{A}+i\sin\gamma\,\mathbf{B}=\cos\gamma(\mathbf{A}+ix\mathbf{B}),$$

where the real and imaginary parts are separated and $x=\tan\gamma$. By actual multiplication of matrices, and especally observing that

$$\mathbf{B}\mathbf{S}(\rho)=\mathbf{S}(-\rho)\mathbf{B}$$

and

$$\mathbf{S}(\rho_j)\mathbf{S}(\rho_k)=\mathbf{S}(\rho_j+\rho_k),$$

we obtain the following results:

$$\mathbf{M}^{(1)}=\cos\gamma_1[\mathbf{A}+ix_1\mathbf{S}(2\rho_1)\mathbf{B}], \quad (6a)$$

$$\mathbf{M}^{(2)}=\cos\gamma_2\cos\gamma_1\{[1-x_2x_1\mathbf{S}(2\rho_2-2\rho_1)]\mathbf{A}+i[x_1\mathbf{S}(2\rho_1)+x_2\mathbf{S}(2\rho_2)]\mathbf{B}\}$$
$$=\left(\prod_{s=1}^{2}\cos\gamma_s\right)\left\{[1-x_2x_1\mathbf{S}(2\rho_2-2\rho_1)]\mathbf{A}+i\left[\sum_{j=1}^{2}x_j\mathbf{S}(2\rho_j)\right]\mathbf{B}\right\}. \quad (6b)$$

$$M^{(3)}=\cos\gamma_3\cos\gamma_2\cos\gamma_1\{[1-x_2x_1\mathbf{S}(2\rho_2-2\rho_1)-x_3x_1\mathbf{S}(2\rho_3-2\rho_1)-x_3x_2\mathbf{S}(2\rho_3-2\rho_2)]\mathbf{A}$$
$$+i[x_1\mathbf{S}(2\rho_1)+x_2\mathbf{S}(2\rho_2)+x_3\mathbf{S}(2\rho_3)-x_3x_2x_1\mathbf{S}(2\rho_3-2\rho_2+2\rho_1)]\mathbf{B}\},$$
$$=\left(\prod_{s=1}^{3}\cos\gamma_s\right)\left\{\left[1-\sum_{k>j=1}^{3}x_kx_j\mathbf{S}(2\rho_k-2\rho_j)\right]\mathbf{A}+i\left[\sum_{j=1}^{3}x_j\mathbf{S}(2\rho_j)-x_3x_2x_1\mathbf{S}(2\rho_3-2\rho_2+2\rho_1)\right]\mathbf{B}\right\}. \quad (6c)$$

$$\mathbf{M}^{(4)}=\cos\gamma_4\cos\gamma_3\cos\gamma_2\cos\gamma_1\{[1-x_2x_1\mathbf{S}(2\rho_2-2\rho_1)-x_3x_1\mathbf{S}(2\rho_3-2\rho_1)-x_4x_1\mathbf{S}(2\rho_4-2\rho_1)$$
$$-x_3x_2\mathbf{S}(2\rho_3-2\rho_2)-x_4x_2\mathbf{S}(2\rho_4-2\rho_2)-x_4x_3\mathbf{S}(2\rho_4-2\rho_3)+x_4x_3x_2x_1\mathbf{S}(2\rho_4-2\rho_3+2\rho_2-2\rho_1)]\mathbf{A}$$
$$+i[x_1\mathbf{S}(2\rho_1)+x_2\mathbf{S}(2\rho_2)+x_3\mathbf{S}(2\rho_3)+x_4\mathbf{S}(2\rho_4)-x_3x_2x_1\mathbf{S}(2\rho_3-2\rho_2+2\rho_1)$$
$$-x_4x_2x_1\mathbf{S}(2\rho_4-2\rho_2+2\rho_1)-x_4x_3x_1\mathbf{S}(2\rho_4-2\rho_3+2\rho_1)-x_4x_3x_2\mathbf{S}(2\rho_4-2\rho_3+2\rho_2)]\mathbf{B}\},$$
$$=\left(\prod_{s=1}^{4}\cos\gamma_s\right)\left\{\left[1-\sum_{k>j=1}^{4}x_kx_j\mathbf{S}(2\rho_k-2\rho_j)+x_4x_3x_2x_1\mathbf{S}(2\rho_4-2\rho_3+2\rho_2-2\rho_1)\right]\mathbf{A}\right.$$
$$\left.+i\left[\sum_{j=1}^{4}x_j\mathbf{S}(2\rho_j)-\sum_{l>k>j=1}^{4}x_lx_kx_j\mathbf{S}(2\rho_l-2\rho_k+2\rho_j)\right]\mathbf{B}\right\}. \quad (6d)$$

We now propose that the transformation matrix $\mathbf{M}^{(n)}$ may be expressed in the following "summed" forms depending on whether n is even or odd:

I. When n is even, say 2ν:

$$\mathbf{M}^{(2\nu)}=\left(\prod_{s=1}^{2\nu}\cos\gamma_s\right)\left\{\left[\sum_{l=0}^{\nu}(-)^l\sum_{j_{2l}>\cdots>j_1=1}^{2\nu}\prod_{k=1}^{2l}x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l}(-1)^k\rho_{j_k}\right)\right]\mathbf{A}\right.$$

$$\left.+i\left[\sum_{l=0}^{\nu-1}(-)^l\sum_{j_{2l+1}>\cdots>j_1=1}^{2\nu}\prod_{k=1}^{2l+1}x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\right]\mathbf{B}\right\}\equiv\left(\prod_{s=1}^{2\nu}\cos\gamma_s\right)\mathbf{N}^{(2\nu)}. \quad (7a)$$

II. When n is odd, say $2\nu+1$:

$$\mathbf{M}^{(2\nu+1)}=\left(\prod_{s=1}^{2\nu+1}\cos\gamma_s\right)\left\{\left[\sum_{l=0}^{\nu}(-)^l\sum_{j_{2l}>\cdots>j_1=1}^{2\nu+1}\prod_{k=1}^{2l}x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l}(-)^k\rho_{j_k}\right)\right]\mathbf{A}\right.$$

$$\left.+i\left[\sum_{l=0}^{\nu}(-)^l\sum_{j_{2l+1}>\cdots>j_1=1}^{2\nu+1}\prod_{k=1}^{2l+1}x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\right]\mathbf{B}\right\}\equiv\left(\prod_{s=1}^{2\nu+1}\cos\gamma_s\right)\mathbf{N}^{(2\nu+1)}. \quad (7b)$$

Here since $l=0$, j_{2l} becomes j_0 which is undefined, we may replace this term by 1. We shall prove the above proposition by mathematical induction. It is easy to see that the formulae (7a) and (7b) hold when $n=1, 2, 3, 4$. We assume that they hold for a certain n and conclude from actual multiplication that they are true when n is replaced by $n+1$.

There are two cases. Let us first consider a certain n which is an even integer, say 2ν. We have to show by induction that the formula (7b) holds when we grant (7a) for $2\nu=n$. By fore multiplying the matrix $\mathbf{M}^{(2\nu)}$ with $\mathbf{M}_{2\nu+1}$, we obtain

$$\mathbf{M}^{(2\nu+1)}\equiv\mathbf{M}_{2\nu+1}\mathbf{M}^{(2\nu)}=\left(\prod_{s=1}^{2\nu+1}\cos\gamma_s\right)\{\mathbf{A}+ix_{2\nu+1}\mathbf{S}(2\rho_{2\nu+1})\mathbf{B}\}\mathbf{N}^{(2\nu)}=\left(\prod_{s=1}^{2\nu+1}\cos\gamma_s\right)\{\mathbf{R}+i\mathbf{I}\},$$

where

$$\mathbf{R}\equiv\left\{\left[\sum_{l=0}^{\nu}(-)^l\sum_{j_{2l}>\cdots>j_1=1}^{2\nu}\prod_{k=1}^{2l}x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l}(-)^k\rho_{j_k}\right)\right]\right.$$

$$\left.+\left[\sum_{l=1}^{\nu}(-)^l\sum_{j_{2l-1}>\cdots>j_1=1}^{2\nu}(x_{2\nu+1})\prod_{k=1}^{2l-1}x_{j_k}\mathbf{S}\left(2\rho_{2\nu+1}-2\sum_{k=1}^{2l-1}(-)^{k-1}\rho_{j_k}\right)\right]\right\}\mathbf{A} \quad (7')$$

and

$$\mathbf{I}\equiv\left\{\left[\sum_{l=0}^{\nu-1}(-)^l\sum_{j_{2l+1}>\cdots>j_1=1}^{2\nu}\prod_{k=1}^{2l+1}x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\right]\right.$$

$$\left.+\left[\sum_{l=0}^{\nu}(-)^l\sum_{j_{2l}>\cdots>j_1=1}^{2\nu}(x_{2\nu+1})\prod_{k=1}^{2l}x_{j_k}\mathbf{S}\left(2\rho_{2\nu+1}-2\sum_{k=1}^{2l}(-)^k\rho_{j_k}\right)\right]\right\}\mathbf{B}. \quad (7'')$$

We note in (7′) that for $l=0$ the sum in the first square bracket will give 1. For any fixed integer l, say λ, greater than zero, we have the sum

$$(-)^\lambda\left\{\left[\sum_{j_{2\lambda}>\cdots>j_1=1}^{2\nu}x_{j_{2\lambda}}\mathbf{S}((-)^{2\lambda}2\rho_{j_{2\lambda}})\cdots x_{j_1}\mathbf{S}(-2\rho_{j_1})\right]\right.$$

$$\left.+\left[x_{2\nu+1}\mathbf{S}(2\rho_{2\nu+1})\sum_{j_{2\lambda-1}>\cdots>j_1=1}^{2\nu}x_{j_{2\lambda-1}}\mathbf{S}((-)^{2\lambda-1}2\rho_{j_{2\lambda-1}})\cdots x_{j_1}\mathbf{S}(-2\rho_{j_1})\right]\right\}\mathbf{A}$$

$$=(-)^\lambda\left\{\left[\sum_{2\nu}\mathbf{a}_{2\lambda}\mathbf{a}_{2\lambda-1}\cdots\mathbf{a}_1\right]+x_{2\nu+1}\mathbf{S}(2\rho_{2\nu+1})\left[\sum_{2\nu}\mathbf{a}_{2\lambda-1}\mathbf{a}_{2\lambda-2}\cdots\mathbf{a}_1\right]\right\}\mathbf{A}$$

$$=(-)^\lambda\left\{\sum_{2\nu+1}\mathbf{a}_{2\lambda}\cdots\mathbf{a}_1\right\}\mathbf{A},$$

where

$$\mathbf{a}_k \equiv x_{j_k}\mathbf{S}((-)^k 2\rho_{j_k}) \quad \text{and} \quad \sum_N \mathbf{a}_t\mathbf{a}_{t-1}\cdots\mathbf{a}_1$$

is an elementary symmetrical function of matrices. The last member of the above equality gives

$$(-)^\lambda\left\{\sum_{j_{2\lambda}>\cdots>j_1=1}^{2\nu+1}\prod_{k=1}^{2\lambda} x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2\lambda}(-)^k\rho_{j_k}\right)\right\}\mathbf{A},$$

a sum in $\mathbf{M}^{(2\nu+1)}$ of (7b).

We shall further note in (7″) that for $l=\nu$ the sum in the second square bracket yields a single term

$$(-)^\nu\{x_{2\nu+1}x_{2\nu}\cdots x_1\mathbf{S}(2\rho_{2\nu+1}-2\rho_{2\nu}+\cdots+2\rho_1)\}\mathbf{B}$$

which is the last term in $\mathbf{M}^{(2\nu+1)}$. For any fixed $l=\lambda$, $0\leqq\lambda<\nu$, and put

$$\mathbf{b}_k \equiv x_{j_k}\mathbf{S}((-)^{k-1}2\rho_{j_k}),$$

we have

$$\begin{aligned}(-)^\lambda\Bigg\{\Bigg[\sum_{j_{2\lambda+1}>\cdots>j_1=1}^{2\nu} & x_{j_{2\lambda+1}}\mathbf{S}((-)^{2\lambda}2\rho_{j_{2\lambda+1}})x_{j_{2\lambda}}\mathbf{S}((-)^{2\lambda-1}2\rho_{j_{2\lambda}})\cdots x_{j_1}\mathbf{S}(2\rho_{j_1})\Bigg] \\ &+\Bigg[x_{2\nu+1}\mathbf{S}(2\rho_{2\nu+1})\sum_{j_{2\lambda}>\cdots>j_1=1}^{2\nu} x_{j_{2\lambda}}\mathbf{S}((-)^{2\lambda-1}2\rho_{j_{2\lambda}})\cdots x_{j_1}\mathbf{S}(2\rho_{j_1})\Bigg]\Bigg\}\mathbf{B} \\ =(-)^\lambda & \left\{\left[\sum_{2\nu}\mathbf{b}_{2\lambda+1}\mathbf{b}_{2\lambda}\cdots\mathbf{b}_1\right]+\left[x_{2\nu+1}\mathbf{S}(2\rho_{2\nu+1})\sum_{2\nu}\mathbf{b}_{2\lambda}\mathbf{b}_{2\lambda-1}\cdots\mathbf{b}_1\right]\right\}\mathbf{B} \\ =(-)^\lambda & \left\{\sum_{2\nu+1}\mathbf{b}_{2\lambda+1}\mathbf{b}_{2\lambda}\cdots\mathbf{b}_1\right\}\mathbf{B} \\ =(-)^\lambda & \left\{\sum_{j_{2\lambda+1}>\cdots>j_1=1}^{2\nu+1}\prod_{k=1}^{2\lambda+1} x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2\lambda+1}(-)^{k-1}\rho_{j_k}\right)\right\}\mathbf{B},\end{aligned}$$

a sum of terms in $\mathbf{M}^{(2\nu+1)}$. Hence we have proved the first case.

The second case, namely, for a certain n which is an odd integer, can similarly be proved. Thus, formulae (7a) and (7b) are completely established.

Furthermore, by comparing (7a) and (7b), we may also combine them into one compact form:

$$\begin{aligned}\mathbf{M}^{(n)}=\left(\prod_{s=1}^{n}\cos\gamma_s\right)\sum_{l=0}^{[n/2]}(-)^l\Bigg\{ & \sum_{j_{2l}>\cdots>j_1=1}^{n}\prod_{k=1}^{2l} x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l}(-)^k\rho_{j_k}\right)\mathbf{A} \\ & +i\sum_{j_{2l+1}>\cdots>j_1=1}^{n}\prod_{k=1}^{2l+1} x_{j_k}\mathbf{S}\left(2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\mathbf{B}\Bigg\}, \quad (7)\end{aligned}$$

where $[l]$ denotes the greatest integer not greater than l, and with the understanding that the term containing j_m where $m>n$ is replaced by zero.

INTENSITY FORMULA

Let us now consider the plane light wave after it passes through the polarizer. Its electric vector $\mathcal{E}_0$ is easily seen to be of the following form:

$$\mathcal{E}_0 \equiv \begin{pmatrix}E_{x0}\\ E_{y0}\end{pmatrix} = \begin{pmatrix}1\\ 0\end{pmatrix}ae^{i\omega t}, \quad (8)$$

where a is the amplitude of the wave. Its intensity I_0 may be calculated as follows:

$$I_0 = K\langle \mathcal{E}_0^* \rangle \mathcal{E}_0 = K(1,0) E_{x_0}^* \begin{pmatrix} 1 \\ 0 \end{pmatrix} E_{x_0} = K E_{x_0}^* E_{x_0} = K a^2, \tag{9}$$

where $\langle \mathcal{E}^* \rangle$ is the conjugate transpose of $\mathcal{E}$, and K the proportionality constant. We have just seen that the electric vector after passing through n retardation plates is

$$\mathcal{E}_n = \mathbf{M}^{(n)} \mathcal{E}_0$$

which has been expressed with respect to the direction of vibration of the polarizer as the x axis. Referring to the axis (a) of the analyzer (Fig. 2) as the new x axis $\mathcal{E}_n$ will become $\mathcal{E}_{n(a)}$, where

$$\mathcal{E}_{n(a)} = \mathbf{S}(-\phi)\mathcal{E}_n = \mathbf{S}(-\phi)\mathbf{M}^{(n)}\mathcal{E}_0.$$

Here ϕ as defined above (Fig. 2) is the orientation of the analyzer with respect to the polarizer p. Now, the electric vector $\mathcal{E}_a$ emerging from the analyzer is plainly the component of the vector relating to the "fast" direction of vibration within the analyzer, thus,

$$\mathcal{E}_a \equiv \begin{pmatrix} E_{x,\,n(a)} \\ 0 \end{pmatrix},$$

where

$$E_{x,\,n(a)} = (1,0)\mathbf{S}(-\phi)\mathbf{M}^{(n)}\mathcal{E}_0.$$

We can easily see that

$$\mathbf{A}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathbf{B}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad (1,0)\mathbf{S}(\rho)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \cos\rho.$$

Hence

$$E_{x,\,n(a)} = \left(\prod_{s=1}^{n} \cos\gamma_s\right)\left\{\sum_{l=0}^{[n/2]} (-)^l \left[\sum_{j_{2l}>\cdots>j_1=1}^{n} \prod_{k=1}^{2l} x_{j_k} \cos\left(\phi - 2\sum_{k=1}^{2l}(-)^k \rho_{j_k}\right)\right.\right.$$

$$\left.\left. + i \sum_{j_{2l+1}>\cdots>j_1=1}^{n} \prod_{k=1}^{2l+1} x_{j_k} \cos\left(\phi - 2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\right]\right\} a e^{i\omega t}. \tag{10}$$

Here we should note that when $l=0$, then $j_{2l}=j_0$, and hence this term is replaced by $\cos\phi$. The intensity from the analyzer is then given by

$$I_n = kE_{x,\,n(a)}^* E_{x,\,n(a)}$$

$$= I_0\left\{\prod_{s=1}^{n}\cos\gamma_s\right\}^2\left\{\left[\sum_{l=0}^{[n/2]}(-)^l \sum_{j_{2l}>\cdots>j_1=1}^{n}\prod_{k=1}^{2l} x_{j_k}\cos\left(\phi - 2\sum_{k=1}^{2l}(-)^k\rho_{j_k}\right)\right]^2\right.$$

$$\left. + \left[\sum_{l=0}^{[n/2]}(-)^l \sum_{j_{2l+1}>\cdots>j_1=1}^{n}\prod_{k=1}^{2l+1} x_{j_k}\cos\left(\phi - 2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\right]^2\right\}, \tag{11}$$

which may be called a generalized formula for the intensity of a system of n retardation plates.

In particular, when $n=3$, we have

$$I_3 = I_0\left\{\prod_{s=1}^{3}\cos\gamma_s\right\}^2\left\{\left[\cos\phi - \sum_{j_2>j_1=1}^{3}\prod_{k=1}^{2} x_{j_k}\cos\left(\phi - 2\sum_{k=1}^{2}(-)^k\rho_{j_k}\right)\right]^2\right.$$

$$\left. + \left[\sum_{j_1=1}^{3} x_{j_1}\cos(\phi - 2\rho_{j_1}) - x_3x_2x_1\cos(\phi - 2\rho_3 + 2\rho_2 - 2\rho_1)\right]^2\right\} \tag{12}$$

or

$$\frac{I_3}{I_0}=[\cos\gamma_3\cos\gamma_2\cos\gamma_1\cos\phi-\{\cos\gamma_3\sin\gamma_2\sin\gamma_1\cos(\phi-2\rho_2+2\rho_1)$$
$$+\cos\gamma_2\sin\gamma_3\sin\gamma_1\cos(\phi-2\rho_3+2\rho_1)+\cos\gamma_1\sin\gamma_3\sin\gamma_2\cos(\phi-2\rho_3+2\rho_2)\}]^2$$
$$+[\{\cos\gamma_3\cos\gamma_2\sin\gamma_1\cos(\phi-2\rho_1)+\cos\gamma_3\cos\gamma_1\sin\gamma_2\cos(\phi-2\rho_2)$$
$$+\cos\gamma_2\cos\gamma_1\sin\gamma_3\cos(\phi-2\rho_3)\}-\sin\gamma_3\sin\gamma_2\sin\gamma_1\cos(\phi-2\rho_3+2\rho_2-2\rho_1)]^2. \quad (12')$$

Similarly, when $n=4$, we get

$$I_4=I_0\left\{\prod_{s=1}^{4}\cos\gamma_s\right\}^2\left\{\left[\cos\phi-\sum_{j_2>j_1=1}^{4}x_{j_2}x_{j_1}\cos(\phi-2\rho_{j_2}+2\rho_{j_1})+x_4x_3x_2x_1\cos(\phi-2\rho_4+2\rho_3-2\rho_2+2\rho_1)\right]^2\right.$$
$$\left.+\left[\sum_{j_1=1}^{4}x_{j_1}\cos(\phi-2\rho_{j_1})-\sum_{j_3>j_2>j_1=1}^{4}x_{j_3}x_{j_2}x_{j_1}\cos(\phi-2\rho_{j_3}+2\rho_{j_2}-2\rho_{j_1})\right]^2\right\} \quad (13)$$

or

$$\frac{I_4}{I_0}=[\cos\gamma_4\cos\gamma_3\cos\gamma_2\cos\gamma_1\cos\phi-\{\cos\gamma_4\cos\gamma_3\sin\gamma_2\sin\gamma_1\cos(\phi-2\rho_2+2\rho_1)$$
$$+\cos\gamma_4\cos\gamma_2\sin\gamma_3\sin\gamma_1\cos(\phi-2\rho_3+2\rho_1)+\cos\gamma_3\cos\gamma_2\sin\gamma_4\sin\gamma_1\cos(\phi-2\rho_4+2\rho_1)$$
$$+\cos\gamma_4\cos\gamma_1\sin\gamma_3\sin\gamma_2\cos(\phi-2\rho_3+2\rho_2)+\cos\gamma_3\cos\gamma_1\sin\gamma_4\sin\gamma_2\cos(\phi-2\rho_4+2\rho_2)$$
$$+\cos\gamma_2\cos\gamma_1\sin\gamma_4\sin\gamma_3\cos(\phi-2\rho_4+2\rho_3)\}$$
$$+\sin\gamma_4\sin\gamma_3\sin\gamma_2\sin\gamma_1\cos(\phi-2\rho_4+2\rho_3-2\rho_2+2\rho_1)]^2$$
$$+[\{\cos\gamma_4\cos\gamma_3\cos\gamma_2\sin\gamma_1\cos(\phi-2\rho_1)+\cos\gamma_4\cos\gamma_3\cos\gamma_1\sin\gamma_2\cos(\phi-2\rho_2)$$
$$+\cos\gamma_4\cos\gamma_2\cos\gamma_1\sin\gamma_3\cos(\phi-2\rho_3)+\cos\gamma_3\cos\gamma_2\cos\gamma_1\sin\gamma_4\cos(\phi-2\rho_4)\}$$
$$-\{\cos\gamma_4\sin\gamma_3\sin\gamma_2\sin\gamma_1\cos(\phi-2\rho_3+2\rho_2-2\rho_1)$$
$$+\cos\gamma_3\sin\gamma_4\sin\gamma_2\sin\gamma_1\cos(\phi-2\rho_4+2\rho_2-2\rho_1)$$
$$+\cos\gamma_2\sin\gamma_4\sin\gamma_3\sin\gamma_1\cos(\phi-2\rho_4+2\rho_3-2\rho_1)$$
$$+\cos\gamma_1\sin\gamma_4\sin\gamma_3\sin\gamma_2\cos(\phi-2\rho_4+2\rho_3-2\rho_2)\}]^2. \quad (13')$$

The intensity formula for *three* plates has been checked to be equal to the result given by Szivessy and Herzog.[5]

GENERAL APPLICATIONS AND A VERIFICATION OF THE INTENSITY FORMULA FOR FOUR PLATES

The theory of optical compensators takes on particular interest in the intensity formulas. When the experimental arrangement of a compensator requires one plate, usually the half-shade plate, in a certain position, the question arises whether a *change of the order* of the plates affects the intensity. Many discussions about this problem, especially of the well-known Brace compensator, can be found in the literature.

From our generalized formula (11), it follows that two consecutive plates cannot, in general, be interchanged. When the optical system is set in "crossed" position, i.e., $\phi=\pm\pi/2$, or in "parallel" position, i.e., $\phi=0$ or $\phi=\pi$, some of the consequences known from the different formulas for two and three plates may be easily deduced from our formula in the "summed" form. These consequences are: In the case of *two* plates p_1 and p_2 can be interchanged without affecting the intensity I_2.[6] In the case of *three*

[5] Reference 3, Szivessy and Herzog.

[6] Szivessy, *Handbuch der Physik* (1928), Vol. 20, p. 764.

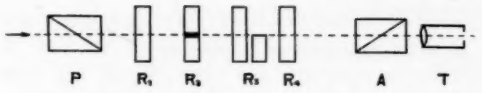

FIG. 3. Optical system of four plates.

plates it is clear from formula (12) that p_1 can be interchanged with p_3, but not p_1 with p_2 or p_2 with p_3.[7] Now, in the case of *four* plates the expanded formula (13′) shows that p_1 with p_4 and p_2 with p_3 must be interchanged simultaneously, if there should be no change of intensity.

All these cases are nothing else than special applications of the general rule: either in crossed or parallel position, if we arrange the system of retardation plates in the *reverse* order but with the same orientations with respect to the "fast" axis of the polarizer, I_n will not be affected. This rule follows readily from (11). For the reversion of the order of the system of plates means mathematically a change of the suffix i of ρ_i and x_i by a substitution $\begin{pmatrix} 1, & 2, & \cdots & n-1, n \\ n, n-1, & & & 2, \quad 1 \end{pmatrix}$. It does not affect the term

$$\left[\sum_{l=0}^{[n/2]}(-)^l \sum_{j_{2l+1}>\cdots>j_1=1}^{n} \prod_{k=1}^{2l+1} x_{j_k} \times\cos\left(\phi-2\sum_{k=1}^{2l+1}(-)^{k-1}\rho_{j_k}\right)\right]^2,$$

while in the term

$$\left[\sum_{l=0}^{[n/2]}(-)^l \sum_{j_{2l}>\cdots>j_1=1}^{n} \prod_{k=1}^{2l} \times\cos\left(\phi-2\sum_{k=1}^{2l}(-)^{k}\rho_{j_k}\right)\right]^2$$

the sum

$$\sum_{k=1}^{2l}(-)^k\rho_{j_k}=\alpha$$

will change its sign. But, for $-\pi<\phi\leqq\pi$ and for any value of α, $\cos(\phi-\alpha)=\cos(\phi+\alpha)$ when and only when $\phi=0$ or $\phi=\pi$, that is the parallel position, and $\cos(\phi-\alpha)=-\cos(\phi+\alpha)$ when and only when $\phi=\pm\pi$, that is the crossed position.

[7] Reference 3, Szivessy and Herzog, p. 54.

In some types of compensators it is admissible to calculate the phase difference by using an approximate intensity formula. Here, the expanded formula (12′) shows that, when the phase differences of the three plates are very small, all terms in γ_i of higher order than the second can be neglected and the three plates can be mutually interchanged for $\phi=\pm\pi/2$.[7] In this case it is also clear from the "summed" formula (11) that any number of plates can be arranged in any order.

Experiments were arranged and carried out to check the above-mentioned statement about the change of order of *four* plates, with small as well as large phase differences.

Furthermore, in order to verify the intensity formula (13′) by well-known experimental results, we have used the following arrangement of a four-plate compensator: let a monochromatic parallel beam of light pass successively through the polarizer P, four retardation plates R_1, R_2, R_3, R_4, the analyzer A and the telescope T focused to plate R_3 (Fig. 3). Plate R_1, a glass plate under tension, acts as a uniaxial crystal whose phase difference $\delta_1=2\gamma_1$ depends upon the tension applied to the glass. Plate R_2 consists of two semicircular quarter-wave plates which divide the light beam into two parts, say I and II. Plate R_3 is a combination of a half-shade plate and the plate whose phase difference $\delta_3'=2\gamma_3'$ may be determined. The half-shade plate was cut and arranged such that only one light beam, say in part II, passed through it. Since the axes of these two plates are purposely taken to be parallel to each other, we shall consider the two plates in part II as one with phase difference $\delta_3''=2\gamma_3''$. R_4 is a quarter-wave plate of the wavelength used. In our experiment the angles of the "fast" directions of vibration, ρ_i, referred to that of the polarizer and the values of the phase differences, $\delta_i=2\gamma_i$, of the plates in Parts I and II of the light beam are specially chosen as shown in Table I. From these given data, we may easily see that the plate system R_1R_2 works like a quarter-wave compensator which produces the rotations ρ and $-\rho$ whose absolute values are equal to γ_1. The system R_3R_4 is a half-shade quarter-wave compensator.

According to the intensity formula (13′) the

TABLE I.

		R_1	R_2	R_3	R_4
I	γ'	γ_1	$\frac{\pi}{4}$	γ_3''	$\frac{\pi}{4}$
	ρ'	$\frac{\pi}{4}$	0	$\frac{\pi}{4}$	0
II	γ''	γ_1	$\frac{\pi}{4}$	γ_3''	$\frac{\pi}{4}$
	ρ''	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{\pi}{4}$	0

intensities I_4' and I_4'' of the light passing through the analyzer are found as

$$I_4' = I_0[\sin^2\gamma_1 \sin^2(\phi+\gamma_3') + \cos^2\gamma_1 \cos^2(\phi+\gamma_3')],$$
$$I_4'' = I_0[\sin^2\gamma_1 \sin^2(\phi+\gamma_3'') + \cos^2\gamma_1 \cos^2(\phi+\gamma_3'')]. \quad (14)$$

The intensity $I=I_4'=I_4''$ for a matched field of view gives the relation between the half-shade azimuths $\bar{\phi}$ of the analyzer and the phase differences δ_1, δ_3', and δ_3'' as

$$(\sin^2\gamma_1-\cos^2\gamma_1)\sin(2\bar{\phi}+\gamma_3'+\gamma_3'')\times\sin(\gamma_3'-\gamma_3'')=0. \quad (15)$$

From this equation follow three possible cases:

(a) $\sin^2\gamma_1-\cos^2\gamma_1=0$, or $\gamma_1=\pi/4$; i.e., R_1 becomes a quarter-wave plate. From the table then it is easily seen that the light beams emerging from R_4 are circularly polarized. It follows directly from (14) that

$$I_4'=I_4''=\tfrac{1}{2}I_0$$

for any azimuth ϕ of the analyzer.

(b) $\sin(\gamma_3'-\gamma_3'')=0$, or $\gamma_3'=\gamma_3''$; i.e. there is no half-shade plate in R_3. The whole system works like a simple quarter-wave compensator, but the intensities of the minima and the maxima depend upon γ_1.

(c) $\sin(2\bar{\phi}+\gamma_3'+\gamma_3'')=0$. The values of the half-shade azimuth $\bar{\phi}$, which are in agreement with the ordinary half-shade compensator, do not depend upon γ_1, but the intensities of these half-shade positions vary with γ_1.

We see in the cases (b) and (c) that the rotations $\pm\rho$ produced by the plate system R_1R_2 do not affect the analyzer azimuths to be measured, but the intensities of these positions. Thus the results of our four-plate compensator are in agreement with the theory of compensators of the analyzer azimuth-measuring type.[8]

[8] M. Richartz, J. Opt. Soc. Am. 31, 292 (1941). Reference 3, Szivessy and Herzog, p. 305; Gabler and Sokob, p. 310.

A New Calculus for the Treatment of Optical Systems

V. A More General Formulation, and Description of Another Calculus

R. CLARK JONES
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts

(Received October 18, 1946)

After a brief review of the nature of the calculus, a formulation is provided for the application of the calculus to a wider class of optical systems than those postulated in the earlier papers of this series. The application is found, however, to be subject to the drastic limitation that the system must not depolarize the light passing through it. Since homogeneous crystalline plates are almost completely free of depolarization, this limitation does not exist in the previously contemplated application to piles of crystalline plates. A more powerful calculus, whose formulation is due to Soleillet and whose development is due to Perrin and Mueller, is described, and is found to be free of this limitation. Its general properties are compared with those of the calculus discussed in this series of papers.

INTRODUCTION

IN 1941 and 1942 there appeared in this Journal a series of four papers[1–4] with the same main title as this one. These four papers, rather heavily burdened with mathematical equations, represented the elaboration of a very simple idea, that of using a two-by-two matrix to represent the effect of a crystalline plate on the state of polarization of the light passing through it.

If one establishes two mutually perpendicular directions, x and y, both of them perpendicular to the direction of propagation of the light, then the polarized light which is incident on the crystal may be described by giving the phases and amplitudes of the x and y components of its electric vector. Both the phases and the amplitudes of these components, denoted by E_x and E_y, may be indicated by permitting E_x and E_y to assume complex values.

It is a physical fact that, for all crystals, the two components E_x' and E_y' of the electric vector of the light emerging from the far side of the crystal are linear functions of the two components of the incident light. Thus one may write:

$$E_x' = m_1 E_x + m_4 E_y, \quad E_y' = m_3 E_x + m_2 E_y, \quad (1)$$

where the m's are constants, in general complex, which are characteristic of the crystal. If one introduces the definitions:

$$\mathcal{E} \equiv \begin{pmatrix} E_x \\ E_y \end{pmatrix}; \quad \mathbf{M} \equiv \begin{pmatrix} m_1 & m_4 \\ m_3 & m_2 \end{pmatrix}, \quad (2)$$

then Eqs. (1) may be written in the simple form:

$$\mathcal{E}' = \mathbf{M}\mathcal{E}. \quad (3)$$

This purely formal step of introducing a matrix notation turned out to be fruitful of new and interesting results, because of the large bag of tricks which may be used in the manipulation of matrices. Nearly all of the operations and concepts which are familiar to those working in

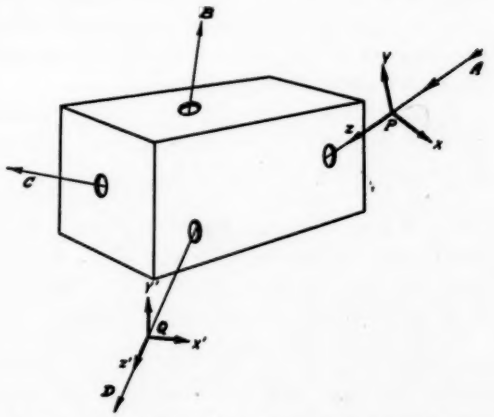

FIG. 1. Showing the box which represents the more general formulation of the calculus contained in this paper. The beam of light A is entering the box, and the beams B, C, and D are emerging from the box. This figure also shows the coordinate systems for the entering and emerging beams described in the text.

[1] I. R. Clark Jones, J. Opt. Soc. Am. **31**, 488–493 (1941).
[2] II. Henry Hurwitz, Jr., and R. Clark Jones, J. Opt. Soc. Am. **31**, 493–499 (1941).
[3] III. R. Clark Jones, J. Opt. Soc. Am. **31**, 500–503 (1941).
[4] IV. R. Clark Jones, J. Opt. Soc. Am. **32**, 486–493 (1942).

quantum mechanics are useful in relation to optical systems: characteristic values and characteristic vectors, coordinate transformation, change of basis vectors, and diagonalization, for example. Indeed the only concept prominent in quantum mechanics that is not useful is that of the Hermitian matrix. Unitary matrices, on the other hand, turn out to be co-extensive with the class of non-absorbing crystals.

A MORE GENERAL FORMULATION

Now it is possible to generalize the physical formulation of the calculus, although in so doing the possibility of a limitation arises which did not exist as long as one was concerned only with homogeneous crystals.

Suppose that the box shown in Fig. 1 represents an optical system—any optical system whatever. Furthermore, suppose that when a beam of light A enters the box, it gives rise to a number of emergent beams, indicated as B, C, D. In general, no two of the beams are supposed to lie in the same plane.

At some point P along the beam A, erect a right-handed set of mutually perpendicular axes x, y, and z, with the z axis parallel to the forward wave normal, and at some point Q along any one of the emergent beams, say D, erect an x', y', and z' coordinate system with z' parallel to the forward wave normal. Place a stop around the beam at P and decrease the maximum diameter of its aperture until the wave front is so small in area that it is essentially flat, and of uniform amplitude and state of polarization.

It is assumed that the light which passes through the aperture in the stop at P is completely polarized and monochromatic.

Then the x' and y' components of the electric vector at Q will be linear functions of the x and y components of the electric vector at P, as indicated by Eqs. (1) and (3), provided that both of the following conditions are satisfied.

Condition **I**. The light at Q is completely polarized—that is to say, the optical system within the box does not depolarize the light passing through it.

Condition **II**. The light at Q is monochromatic and of the same frequency as the light at P. This restriction rules out Raman effect and fluorescence within the optical system.

Thus, subject to these two important conditions, it has been possible to generalize the calculus so that it holds not merely for plane-parallel crystalline plates, but for a much wider class of optical systems. The generalization has been accomplished, however, only at the cost of the simplicity of description which characterizes the application to plane-parallel plates of crystalline material.

The Two Conditions

The second of the two conditions is not important in practice, but the first condition is very important, because nearly all actual optical systems depolarize the light passing through them to some extent.

The first condition exists because a simple statement of the electric vector of the light is unable to express a state of polarization which is not complete. In order to specify a state of partial polarization, it is necessary to specify two electric vectors, with the proviso that the light consists of the addition of the two vectors without any correlation between their phases—that is to say, that they be added incoherently.

The calculus is able to treat the passage of partially polarized light through a non-depolarizing optical system in the manner discussed in IV, but it is completely unable to treat systems which serve to depolarize the light which passes through them.

The restriction of the first condition may be avoided altogether by the use of a more powerful calculus which Professor Hans Mueller has developed in considerable detail in a course of lectures during 1945–1946 at the Massachusetts Institute of Technology.

THE MORE POWERFUL CALCULUS

In order to develop a calculus for the treatment of depolarizing optical systems, it is first of all desirable to have a concise means of representing partially polarized light. Such a representation is provided by the Stokes parameters[5] of a beam of light.

Any beam of partially polarized light may be considered as the incoherent superposition of two beams, the first of which is natural (unpolar-

[5] G. G. Stokes, Trans. Camb. Phil. Soc. 9, 399 (1852).

ized) light, and the second of which is completely polarized light. The resolution into the two beams is unique.

The beam of natural light is characterized completely by the r.m.s. amplitude A of its electric vector. The completely polarized beam may be defined by the r.m.s. amplitudes A_x and A_y of the x and y components of the electric vector, and by their relative phase ϕ.

The Stokes parameters of the beam are defined by

$$\begin{aligned} I &\equiv A_x^2+A_y^2+A^2, \\ M &\equiv A_x^2-A_y^2, \\ C &\equiv 2A_xA_y\cos\phi, \\ S &\equiv 2A_xA_y\sin\phi, \end{aligned} \tag{4}$$

and may be arranged to form a vector, the Stokes vector $\mathfrak{S}$, in the manner

$$\mathfrak{S} \equiv \begin{pmatrix} I \\ M \\ C \\ S \end{pmatrix}. \tag{5}$$

The Stokes vector has the important property that, if two separate beams of light are superposed incoherently, then the Stokes vector of the combination is the sum of the Stokes vectors of the two original beams. On the other hand, if the two beams are added coherently, the Stokes vector of the combination is still calculable from the Stokes vectors of the individual beams, but is not obtained by simple addition of the vectors.

On the other hand, the $\mathcal{E}$ vectors are added if the two beams are combined coherently, whereas if the two beams are superposed incoherently, the result is not expressible as an $\mathcal{E}$ vector, since the incoherent combination of two beams generally results in partially polarized light.

The Stokes vector and the $\mathcal{E}$ vector both require the specification of four numbers to determine them. In the case of the Stokes vector, the four numbers are the four Stokes parameters, whereas in the case of the $\mathcal{E}$ vector, the four numbers are the real and imaginary parts of the two components. In either of the vectors, three of the four numbers may be considered as specifying the state of polarization of the light. With the Stokes vector, the remaining number may be considered as the specification of the degree of polarization of the light, whereas, with the $\mathcal{E}$ vector, the degree of polarization is always 100 percent, and the fourth number determines the absolute phase of the light.

In the case of completely polarized light, the Stokes vector can be calculated from the $\mathcal{E}$ vector, but, if the Stokes vector is given, the absolute phase of the $\mathcal{E}$ vector is left undetermined.

The Mueller Matrices

It has been shown by Soleillet[6] that the four components of the vector $\mathfrak{S}'$ which represents the light emerging from any optical system which satisfies condition **II** are linear functions of the four components of the vector $\mathfrak{S}$ which represents the light entering the optical system. Thus one may write

$$\mathfrak{S}' = \boldsymbol{M}\mathfrak{S}, \tag{6}$$

where $\boldsymbol{M}$, a matrix introduced by Mueller, is a four-by-four matrix with 16 real matrix elements.

Thus the Mueller matrices permit one to treat optical systems which depolarize light in a way which is very similar to that in which the two-by-two matrices permit one to treat non-depolarizing systems.

As already mentioned, Professor Mueller has elaborated the use of these four-by-four matrices in a recent course of lectures.[7] He expects to include a comprehensive treatment of their properties and applications in a forthcoming book on polarized light.

Comparison

The Mueller matrices contain 16 coefficients, whereas the two-by-two matrices contain 8 parameters (real and imaginary part of each of the four matrix elements). The two-by-two matrices, however, specify the absolute phase of the optical system, whereas the Mueller matrices do not. Thus the significant comparison is 8 compared with 17 coefficients if the absolute phase is included, or 7 compared with 16 if it is not.

If the optical system does not depolarize, then

---

[6] Paul Soleillet, Ann. de physique **12**, 23–97 (1929). See also Francis Perrin, J. Chem. Phys. **10**, 415–427 (1942).

[7] The Mueller matrices are used in a declassified report by Hans Mueller, "Memorandum on the polarization optics of the photo-elastic shutter," Report No. 2 of OSRD, project OEMsr-576, November 15, 1943.

9 identities exist among the 16 Mueller coefficients, so that only 7 of them are independent. Under this condition, the Mueller matrix may be computed from the two-by-two matrix, and the two-by-two matrix, except for a phase factor, may be computed from the Mueller matrix. If the optical system does depolarize, however, none of the 9 identities holds in the most general case.

The primary usefulness of both of the calculi lies in treating light which passes through the successive elements of the optical system in a prescribed way. The existence of multiple paths, particularly those provided by interreflections, increases the complexity of the calculation. When multiple paths do exist, the choice of the calculus may depend on whether the multiple beams are combined coherently, or incoherently.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 2 FEBRUARY, 1947

A New Calculus for the Treatment of Optical Systems

VI. Experimental Determination of the Matrix*

R. CLARK JONES
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts
(Received October 18, 1946)

A simple, straightforward method is described for determining experimentally the matrix of any crystalline plate. The method involves three measurements of the state of polarization of light transmitted by the plate, and also a measurement of its transmission factor for natural light. These measurements determine the matrix uniquely, except for a phase factor whose practical significance is small. The conditions are stated for the application of the method to the more general type of optical system described in V. A statement of the content of future papers in this series is included.

THE PROBLEM

LET us examine the following problem: Suppose that someone has presented us with a plane-parallel slab of crystalline material, and that he wishes to know the two-by-two matrix which represents the crystal for normal incidence at a given wave-length. He has obligingly scratched on one of the two plane surfaces the x and y coordinate axes with respect to which he wishes the matrix stated, and he has specified on which side the incident light is to fall.

A SIMPLE SOLUTION

There are many ways by which the measurement may be done, but the following is probably one of the most simple. The measurement is made in five steps.

1. Let the incident light be linearly polarized with its electric vector parallel to the x axis of the slab, so that E_y is zero. Then the fundamental pair of equations (Eqs. (1) of V)

$$E_x' = m_1 E_x + m_4 E_y, \quad E_y' = m_3 E_x + m_2 E_y, \quad (1)$$

become for this special case:

$$E_x' = m_1 E_x, \quad E_y' = m_3 E_x, \quad (2)$$

or

$$m_1/m_3 = E_x'/E_y', \quad (3)$$

where, in accordance with conditions **I** and **II** of the preceding paper, it is assumed that the emergent light is monochromatic and completely polarized. Now the measurement of the complex ratio E_x'/E_y' is equivalent to the measurement of the state of the polarization of the emergent light, which may be accomplished in a routine manner with a quartz wedge. Thus one may determine the complex ratio $k_1 \equiv m_1/m_3$.

The determination of the state of polarization of the emergent light involves the determination of the axial ratio and sense of rotation of the vibration ellipse, and the determination of the

* Presented at the Thirty-First Annual Meeting of the Optical Society of America, October 3–5, 1946, New York, New York.

azimuth of the major axis of this ellipse. By Eqs. (27) and (28) of I, these determinations may be used to calculate the ratio r of the amplitude of the x component to the amplitude of the y component of the electric vector of the emergent light, and to calculate the phase ϕ of the x component relative to that of the y component. In terms of r and ϕ, the ratio k_1 defined in the last paragraph is equal to $re^{i\phi}$.

2. The measurement is repeated with the electric vector of the incident light parallel to the y axis of the slab. This determines $k_2 \equiv m_4/m_2$.

The slab is now turned over, so that the light passes through it in the opposite direction. If one continues to use the x and y axes which are scratched on the slab, the matrix to be used for the reverse direction of transmission is the transpose of the matrix for the forward direction, so that the equations which replace (1) are:

$$E_x' = m_1 E_x + m_3 E_y, \quad E_y' = m_4 E_x + m_2 E_y. \tag{4}$$

3. With the electric vector of the incident light parallel to the x axis of the inverted slab, one now determines $k_3 \equiv m_1/m_4$.

4. With the electric vector parallel to the y axis, one finally determines $k_4 \equiv m_3/m_2$.

Actually, only three of the four measurements indicated above are independent, since one has

$$k_1 k_4 \equiv k_2 k_3. \tag{5}$$

The existence of this relation does serve, however, as a check on the internal consistency of the measurements. In fact, this relation really involves two independent checks, since both the real and the imaginary parts of the two products must be equal.

By solving for the m's in terms of the k's, one finds that the matrix $\mathbf{M}$ for the forward direction of transmission is

$$\mathbf{M} = A \begin{pmatrix} k_1 k_4 & k_2 \\ k_4 & 1 \end{pmatrix}, \tag{6}$$

where A is a complex constant which is so far undetermined.

The absolute value of A determines the transmission factor of the slab, and is accordingly determined by any single measurement of the transmission factor. From the point of view of simplicity and symmetry, perhaps the most attractive measurement to make is that of the transmission factor of the plate for unpolarized (natural) light. It has been shown in IV that this transmission factor is given by

$$T_{\text{nat}} = \tfrac{1}{2}[|m_1|^2 + |m_2|^2 + |m_3|^2 + |m_4|^2], \tag{7}$$

so that the absolute value of A in terms of T_{nat} and the k's is

$$|A|^2 = \frac{2T_{\text{nat}}}{1 + |k_2|^2 + |k_4|^2 + |k_1|^2 |k_4|^2}. \tag{8}$$

5. Thus the absolute value of A may be determined by a measurement of the transmission factor of the slab for natural light.

Actually, however, because of the practical difficulty of obtaining completely unpolarized monochromatic light, it will be more convenient in practice to measure the transmission factors for the incident light which is used in either step **1** or **2**. One easily finds from (6) that the transmission factor for the incident light used in **1** is given by

$$T_1 = |A|^2 |k_4|^2 (1 + |k_1|^2), \tag{9}$$

and the corresponding transmission factor for step **2**

$$T_2 = |A|^2 (1 + |k_2|^2), \tag{10}$$

whence $|A|$ is given by either of the following two expressions

$$\begin{aligned} |A|^2 &= \frac{T_1}{|k_4|^2 (1 + |k_1|^2)} \\ &= \frac{T_2}{1 + |k_2|^2}, \end{aligned} \tag{11}$$

where the quantities on the right are all measured quantities.

This completes the determination of the matrix of the slab for most practical purposes, since the absolute phase of the matrix is only rarely of interest. Throughout the developments in the previous papers of this series, the phase was chosen so that the determinant of the matrix was real and positive. If this is done, the value of A is completely determined.

If, however, one wishes to determine the value of the phase of A, it is necessary to use some method which is in principle capable of measuring the absolute phase retardation of an optical

system. One method is to interpose the slab in one of the two paths of an interferometer, and to determine the shift in the fringe pattern. This measurement must be made with light whose state of polarization is that of one of the two eigenvectors of the slab. The eigenvectors of $\boldsymbol{M}$ may be determined without knowing the value of A because they are independent of the value of A and depend only on the k's.

THE MORE GENERAL PROBLEM

In the case of the more general type of optical system described in V, the same method of solution may be used, provided first that the conditions **I** and **II** are satisfied, so that the calculus is applicable, and provided second that the system satisfies the principle of reciprocity. The principle of reciprocity is assumed in the use of the transpose of **M** as the matrix for the reverse transmission of the light. In practice, any optical system which satisfies conditions **I** and **II** will satisfy the reciprocity principle, provided no magnetic fields are present. Thus the absence of magnetic fields is the primary condition for the applicability of the solution described above to the more general type of optical system.

SOLUTION WHEN RECIPROCITY FAILS

In the event that the optical system does not satisfy the reciprocity principle, it is still possible to obtain a simple solution of the problem. Instead of the measurements numbered **3** and **4**, the following two measurements should be made.

6. With the electric vector of the incident linearly polarized light parallel to the bisector of the angle between the positive x and y axes, so that $E_x=E_y$, one has from Eq. (1)

$$E_x'=(m_1+m_4)E_x, \quad E_y'=(m_3+m_2)E_x. \quad (12)$$

This permits the determination of the complex ratio

$$k_6\equiv\frac{m_1+m_4}{m_3+m_2}.$$

7. Similarly, with the electric vector of the incident light perpendicular to the bisector of the positive x and y axes, so that $E_x=-E_y$, one determines:

$$k_7\equiv\frac{m_1-m_4}{m_3-m_2}.$$

With the aid of a few lines of algebra, one finds that the constant k_4 is related to k_1, k_2, and k_6 or to k_1, k_2, and k_7 by the relations:

$$k_4=\frac{k_6-k_2}{k_1-k_6}=\frac{k_7-k_2}{k_7-k_1}. \quad (13)$$

Just as with the third and fourth measurements, the sixth and seventh measurements are not independent, as indicated by the second equality in Eq. (13).

With the ratio k_4 obtained by Eq. (13), the determination of the matrix is just as described in the previous section. In the case in which reciprocity fails, it is, of course, necessary to repeat measurements **1**, **2**, **5**, and **6** or **7** with the opposite direction of transmission of the light if one wishes to know the matrix for the opposite direction of transmission.

This completes the description of the method used to obtain the matrix when the system does not satisfy the reciprocity principle.

STATEMENT

Future papers in this series will include a discussion of the matrix as a solution of a differential equation, and the relation between the matrix of a crystalline plate and the complex dielectric tensor of the plate.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 37, NUMBER 2 FEBRUARY, 1947

High Transmission Windows for Radiation of 3 to 14 Microns Wave-Length

MARK HYMAN, JR., AND BRUCE H. BILLINGS
Research Laboratories, Polaroid Corporation, Cambridge, Massachusetts
(Received October 28, 1946)

I. INTRODUCTION

THE development of a window of large area for the transmission of radiation over the range 3μ to 14μ is desirable for the broadening of the scope of infra-red optics.

Of the known materials for this range, silver chloride, despite its several disadvantages, appeared to be the most desirable, particularly if some of its disadvantages could be overcome. This paper describes an attempt to improve the performance of silver chloride windows.

The points in favor of a silver chloride window may be listed as follows:

1. Its absorption is negligible in the above range.
2. Its dispersion is very low in the above range.
3. It is readily available either as a powder or in large sheets.
4. It is readily shaped under pressure.

The undesirable characteristics are:

1. It is an extremely soft material, having a Moh hardness of one—comparable to talc.
2. When it is in contact with many of the common metals and alloys, silver is plated out through a single replacement reaction.
3. It transmits only about 80 percent of the incident radiation in the wave range under discussion, the remainder of the radiation being reflected.
4. It breaks down into its component elements when exposed to light of wave-lengths shorter than 450 mμ.

To overcome the first two defects, it is necessary merely to handle the material with proper precautions against any abrasion or contact with metals. To improve upon the last two characteristics is the purpose of this investigation.

II. THEORY OF NON-REFLECTION COATS ON SILVER CHLORIDE

The use of interference films to reduce reflection loss has been well-known for several years.[1-3] However, with one exception,[4] no attempt, to our knowledge, has heretofore been made to apply this reflection-reducing technique to the infra-red region.

The general equation for the reflectivity R of one side of a material coated with a single transparent layer can be derived from electromagnetic theory[5] and is:

$$R=|\rho|^2, \tag{1}$$

where

$$\rho=\frac{r_{21}+r_{10}e^{i\theta}}{1+r_{21}r_{10}e^{i\theta}}. \tag{2}$$

θ is the angular retardation for a double traversal of the coating layer; and i is $(-1)^{\frac{1}{2}}$.

The subscript 0 refers to the material to be coated; the subscript 1 to the coating layer; and the subscript 2 to the medium in which the system is immersed.

$r_{k,k-1}$ is the amplitude reflectivity of the $k, k-1$ interface when the light is incident in the kth layer.

We shall apply Eq. (2) only to light at normal incidence. Fresnel's laws give for this case

$$r_{k,k-1}=\frac{n_k-n_{k-1}}{n_k+n_{k-1}}, \tag{3}$$

where n_k is the refractive index of the kth layer.

For zero reflectivity at a given wave-length R in Eq. (1) is set equal to zero. The solution is then $e^{i\theta}=-1$, or, for the thinnest layer, $\theta=\pi$. Thus the retardation for a single traversal of the coating layer is $\pi/2$, an optical thickness of $\frac{1}{4}$ wave-length. This solution in conjunction with Eq. (3) gives the result $n_1=(n_0n_2)^{\frac{1}{2}}$. Thus the coating layer should have a refractive index equal to the square root of the product of the indices of the base material and the surrounding medium.

With $n_2=1.0$ (index of air) and $n_0=2.0$ (index of silver chloride at 7μ) n_1 becomes 1.41. If a

[1] J. T. Strong, J. Opt. Soc. Am. **26**, 73 (1936).
[2] C. Hawley Cartwright, Bull. Am. Phys. Soc. **13**, 10 (1938).
[3] K. B. Blodgett, J. Phys. Chem. **41**, 975 (1937).
[4] J. T. Strong, in a private communication, describes an experiment he performed along these lines, but the high absorption of his sample of silver chloride precluded the possibility of his reaching high transmission values.
[5] See, e.g., R. L. Mooney, J. Opt. Soc. Am. **35**, 574 (1945).

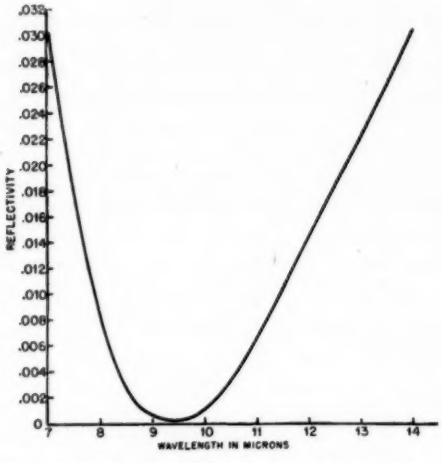

FIG. 1. Theoretical reflectivity of one side of a silver chloride flat coated with a layer of material of index 1.41 and an optical thickness of $1/4\lambda$ at a wave-length of 9.3μ.

material with such an index were applied with an optical thickness of $\frac{1}{4}\lambda$ at 9.3μ, the reflectivity for one side at various wave-lengths would be given by the curve in Fig. 1 and its average value from 7μ to 14μ would be 1.1 percent.

III. ACTUAL COATING USED

After experiments with various materials and methods of their application, it was found that a simple method of forming the non-reflection coat was to pour a solution of a plastic onto the surface of a rapidly rotating silver chloride disk. This results in a coat of uniform thickness, the thickness depending on such variables as the rate of spin, the viscosity of the solution, the type of plastic used, and the vapor pressure of the solvent.

The choice of the particular plastic was determined by two considerations. First, its refractive index should approach as closely as possible to the value of 1.41, which was calculated in the preceding section. Second, it should absorb as little radiation as possible in the 3–14μ range.

Although methyl methacrylate has $n_D=1.49$ and *t*-butyl methacrylate $n_D=1.46$, these are not satisfactory materials because of the presence of the strong absorption band characteristic of esters around 5.8μ and further strong absorption bands in the 8μ to 9μ region.[6] Indeed, all plastics containing oxygen are unsatisfactory due to numerous absorption bands; and the type of plastic indicated is a hydrocarbon—preferably saturated.

Polyethylene, with $n_D=1.51$, immediately comes to mind, but, unfortunately, it is not appreciably soluble at room temperature, and hence cannot be applied readily from solution. The material chosen was hydrogenated polystyrene, (supplied to us in experimental quantities by Dow Chemical Company through the kindness of Dr. R. F. Boyer) which dissolves readily in methyl cyclohexane. This has $n_D=1.51$ and transmits well in all wave-lengths except those corresponding to carbon-hydrogen absorption, namely: 3.4μ–3.5μ and 6.9μ–7.0μ. Naturally, the carbon-hydrogen absorptions are inevitable in all organic plastics.

Figure 2 shows a spectral transmission curve for silver chloride coated on each side with a layer of hydrogenated polystyrene. This particular sample was designed to fit a requirement for peak transmission at 9μ, where it transmits 96 percent, and was also required to have a good transmission in the range 7–13μ, where the average transmission is 93 percent. The optical thickness of this coating is 2.25μ, and its geometrical thickness is about 1.5μ.

IV. THEORY AND PRACTICE OF PROTECTIVE COATING OF STIBNITE ON SILVER CHLORIDE

The photochemical decomposition of silver chloride can occur only insofar as radiation is absorbed by the silver chloride. As shown in Fig. 3, the main absorption of radiation by silver

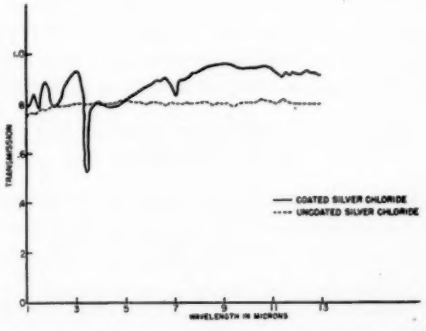

FIG. 2. Transmission of silver chloride flat coated on both sides with hydrogenated polystyrene to an optical thickness of $1/4\lambda$ at 9.0μ.

[6] Barnes, Gore, Liddell, and Williams, *Infra-Red Spectroscopy, Industrial Application, and Bibliography* (Reinhold Publishing Corporation, New York, 1944).

chloride occurs in the blue and ultraviolet regions; hence, any material with close to complete absorption in this region, combined with good transmission in the infra-red region, may be used as a protective coating for the silver chloride. Since the materials which we used for non-reflecting coats do not absorb the blue and ultraviolet, it was necessary to use an additional coating material for protective purposes.

The material found most satisfactory by this laboratory was antimony trisulfide, Sb_2S_3, evaporated at a pressure of 10^{-4} mm of mercury. Before evaporation the trisulfide was in the form of crystals of the mineral stibnite. Although a crystallographic analysis of such evaporated antimony trisulfide films has not yet been made, the films themselves are often referred to as stibnite. The use of this particular material as protection for silver chloride has also been discussed in a report by A. H. Pfund.[7] The reason for the efficacy of this material as a protective agent for the silver chloride is illustrated by Fig. 3, in which is shown the absorption of evaporated stibnite coated onto silver chloride to an optical thickness of $7/4\lambda$ of red light. It absorbs completely below 470 mμ, and hence allows no photochemically active light to penetrate to the silver chloride layer.

The efficacy of this protective coating is shown by Figs. 4–6, which give the results of exposure tests of uncoated silver chloride, $\frac{1}{4}\lambda$ Sb_2S_3 coated silver chloride, $7/4\lambda$ stibnite-coated silver chloride, and $13/4\lambda$ Sb_2S_3 coated silver

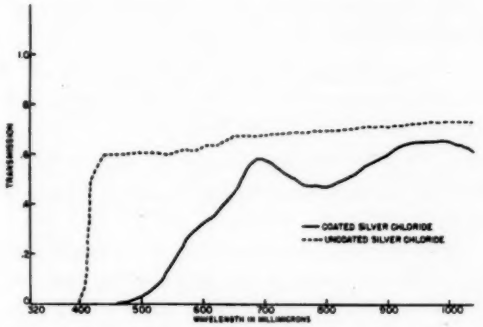

FIG. 3. Transmission of a silver chloride flat coated on one side with evaporated stibnite to an optical thickness of $7/4\lambda$ at a wave-length of 625 mμ.

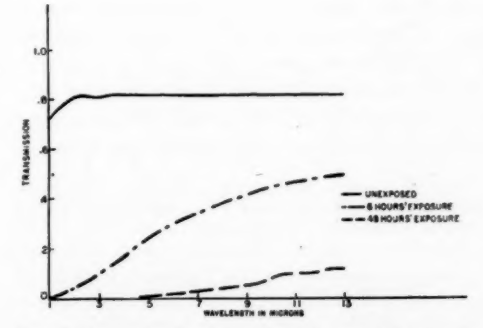

FIG. 4. Transmission of uncoated silver chloride exposed to a General Electric S-1 sunlamp at 7″ distance for various periods.

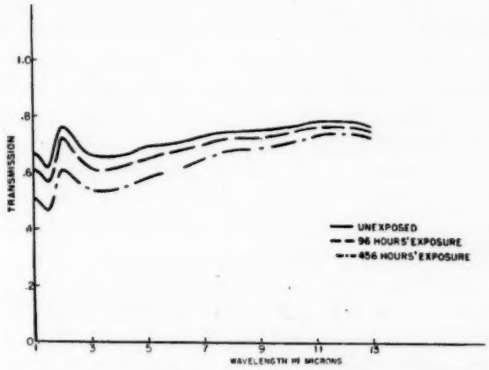

FIG. 5. Transmission of a silver chloride flat coated on one side with a layer of evaporated stibnite of $7/4\lambda$ optical thickness at 625 mμ and exposed for various periods to radiation from a General Electric S-1 sunlamp at 7 inches.

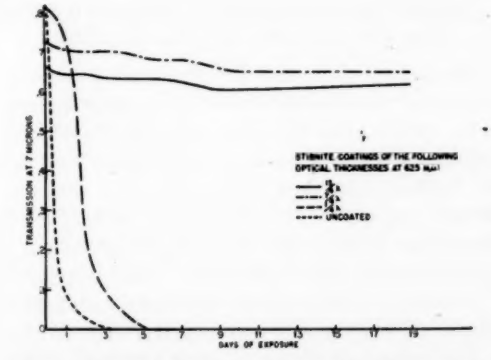

FIG. 6. Transmission at 7μ of silver chloride coated with different thicknesses of evaporated stibnite, as a function of time of exposure to a General Electric S-1 sunlamp at 7″ distance.

[7] A. H. Pfund, OSRD Rep. No. 3552 of April 4, 1944.

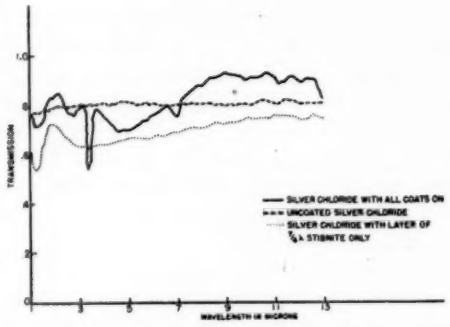

FIG. 7. Transmission of a silver chloride flat—one side coated with evaporated stibnite with an optical thickness of 7/4λ at 625 mμ followed by a layer of polystyrene—the other side coated with hydrogenated polystyrene.

chloride. The exposure in these tests was made at a distance of seven inches from a General Electric S-1 Sunlamp. The results show that after 72 hours continuous exposure to this lamp, unprotected silver chloride has no transmission at 7μ; whereas, after 456 hours continuous exposure, silver chloride with a 7/4λ stibnite coating has dropped in transmission only from 76 percent to 65 percent.

Since the Sb_2S_3 index is higher than the AgCl index, no increase in transmission of AgCl can be attained at any wave-length by application of this coating alone. We may derive from Eq. (1) the conclusion that with this combination the maximum transmission possible is equal to the transmission of the uncoated silver chloride.

V. THEORY AND PRACTICE OF COMBINED STIBNITE AND LOW REFLECTION COATING

Finally, a silver chloride window has been developed, combining protection against actinic light with a low reflection coating. As shown below, a high index plastic is indicated to coat the stibnite-coated window. For this purpose polystyrene ($n_D=1.59$) proved satisfactory. The absorption bands due to the phenyl group's unsaturation, which normally occur at 6.3μ, 6.7μ, and 9.3μ, proved very weak in the thin film employed. On the side not coated with stibnite, a low reflection coat of hydrogenated polystyrene is used. One such combination, designed to give peak transmission at 9μ, is illustrated by the spectral curve of Fig. 7. This curve shows that the low reflection coatings raise the average transmission for 7–13μ from 73 percent to 89 percent.

The treatment of the silver chloride plates, then, is carried out in three steps: (1) The side exposed to visible light is coated with a material which will pass the infra-red and absorb both the shorter wave-length visible light and the ultraviolet. (2) This layer is covered with a material of such thickness and index that it reduces the reflection from this side of the plate to a minimum in the desired spectral region. (3) The unexposed side of the plate is coated with a material of such index and thickness that the reflectivity of this side is reduced to a minimum.

The minimum thickness of the first layer is determined empirically by the probable exposure of the window to visible and ultraviolet light. The question of optimum maximum thickness is of interest and may be determined by the following considerations.

The amplitude reflectivity of the double layer on silver chloride is given by

$$\rho=\frac{r_{32}+r_{21}e^{i\theta_2}+r_{10}e^{i(\theta_2+\theta_1)}+r_{32}r_{21}r_{10}e^{i\theta_1}}{1+r_{21}r_{10}e^{i\theta_1}+r_{32}r_{21}e^{i\theta_2}+r_{10}r_{32}e^{i(\theta_2+\theta_1)}}, \quad (4)$$

where $r_{i,i-1}$ is defined in Eq. (3) and θ_j is the double angular retardation through the jth layer.

There are an infinite number of values for the thickness and index of the two coating layers to give zero reflectivity at one wave-length. We want a value for the thickness and index of the second layer which will make the average reflectivity from 7–14μ as low as possible. The refractive index of the Sb_2S_3 layer is higher than the index of the AgCl. If it is also higher than the index of the second layer then, at a wave-length at which it has an optical thickness of an integral multiple of one half wave, it will have no effect on the reflectivity of the system. At these wave-lengths if the index of the second coat is $(n_0n_3)^{\frac{1}{2}}$,

TABLE I. Average reflectivity of a silver chloride slab coated with material of index 4 followed with a material to reduce reflectivity at 10μ to zero.

Optical thickness of high index layer	Average reflectivity from 7–14μ
0.313μ	2.5%
5.0μ	12.7%
10.0μ	31.1%

and if this coat has an optical thickness of one quarter wave, the reflectivity will be zero.

To have the lowest average reflectivity over a given wave-length range, the thickness of the Sb_2S_3 layer at the frequency of zero reflectivity should be the smallest possible number of half wave-lengths.

To demonstrate this conclusion, a calculation was made of the average 7–14μ reflectivity of a silver chloride flat coated with a material of index 4. The values of the average reflectivity are summarized in Table I.

To get the average reflectivity Eq. (4) was first solved to give the index and thickness of material for zero reflectivity at 10μ. With this value, the reflectivity *vs.* wave-length curve was plotted and integrated with a planimeter. Figure 8 and Fig. 9 illustrate the reflectivity curves of Table I. The process of solving Eq. (4) is shown below.

For the determination of the actual thickness used, the index of the Sb_2S_3 layer was measured at 10μ and found to be approximately 2.5.

As shown above, the Sb_2S_3 optical thickness necessary for protection was only 1μ. With this thickness and the known index at 10μ the only unknowns in Eq. (4) were the index and thickness of the outer layer. Equation (4) could thus be rewritten

$$\rho=\frac{a(\theta, n)+ib(\theta, n)}{c(\theta, n)+id(\theta, n)}, \tag{5}$$

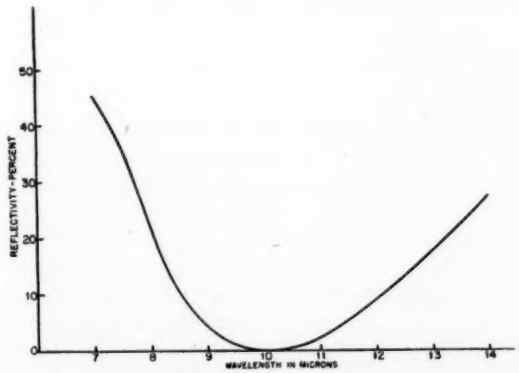

FIG. 8. Theoretical reflectivity of a single surface of silver chloride coated with an optical thickness 5μ of material of index 4 followed by 2.5μ of material of index 1.414.

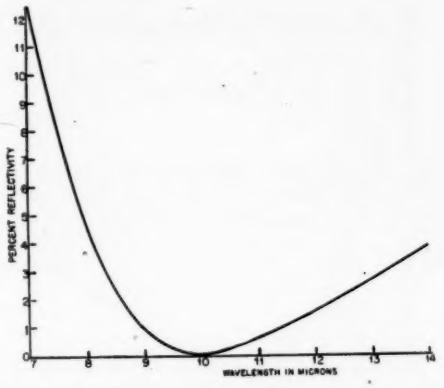

FIG. 9. Calculated reflectivity of a single surface of silver chloride coated with an optical thickness of .313μ of material of index 4 followed by 3.05μ of material of index 1.55.

where n and θ are the index and double angular retardation of the outer layer. When $\rho=0$, it is necessary that $a(\theta, n)=0$, and that $\partial a/\partial\theta=0$. We find

$$a(\theta, n)=A(n)\cos\theta-B(n)\sin\theta-C(n)=0, \tag{6}$$

$$\partial a/\partial\theta=-A(n)\sin\theta-B(n)\cos\theta=0. \tag{7}$$

From these we find

$$\tan\theta=-B(n)/A(n), \tag{8}$$

$$A(n)^2+B(n)^2=C(n)^2. \tag{9}$$

These last two equations were solved for the value of the thickness and index of the protective layer to give zero reflectivity at 10μ. This value and the resulting parameters for the reflection-reducing layer are summarized in Table II.

Reflectivities were calculated for these sets of values and are shown in Fig. 10. This curve is plotted around 10μ as the value λ_0. The 7–14μ average value of the reflectivity is 2.5 percent.

VI. EXPERIMENTAL

The silver chloride sheeting used was 0.072 inch thick and was obtained from the Harshaw Chemical Company.

The stibnite was obtained from the J. T. Baker Chemical Company. It was evaporated in a high vacuum apparatus consisting of an oil-diffusion pump in series with a fore-pump, the pressure during evaporation being about 10^{-4} mm of mercury. The stibnite was placed in a porcelain

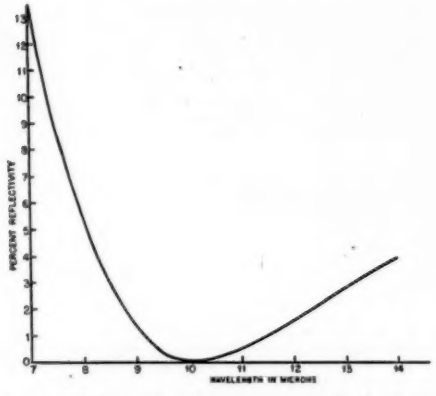

FIG. 10. Calculated reflectivity of a single surface of silver chloride coated with an optical thickness of 1μ of material of index 2.5 followed by 2.86μ of material of index 1.51.

crucible and was heated with a spiral coil of tungsten placed directly over the stibnite surface. The disk of silver chloride, $1\frac{5}{8}$ inches in diameter, was placed about 6 inches above the crucible. The thickness of the coating was controlled by observation of the color change with thickness of coating: when red appeared for the first time, it was coated with $\frac{1}{4}\lambda$ for red light; when red appeared for the second time, the coating was $\frac{3}{4}\lambda$ thick; and for each additional color cycle $\frac{1}{2}\lambda$ was added.

The low reflection coating solutions consisted of hydrogenated polystyrene dissolved in methyl cyclohexane and polystyrene dissolved in isopropyl acetate. These were dropped onto the spinning silver chloride disk in $\frac{1}{2}$ ml quantities from a glass tubing with rubber tubing fastened to one end; the liquid was held in the glass tubing by pinching the rubber, and was released by releasing the pinch. The disk was spun at 1300 rev./min. A constant speed rotator manufactured by the Central Scientific Company (Catalog No. 74350) proved very satisfactory for this purpose when it was modified to include a holder for the disk and a catch-pan for the excess solution.

TABLE II.

Protective layer		Outer layer	
Index at 10μ	Optical thickness	Index at 10μ	Optical thickness
2.5	1.00μ	1.51	2.86μ

Infra-red transmission measurements were made on a Perkin-Elmer Infra-Red Spectrometer, Model 12A.

VII. CONCLUSION

Silver chloride windows for the transmission of infra-red have been protected against photochemical decomposition by coating with evaporated stibnite. They have, further, been rendered high transmitting by coating with various plastics.

The authors wish to acknowledge with thanks the assistance of several members of the Polaroid Laboratories and particularly that of Miss Emma Comerford in the calculations and that of Mr. Edgar Barr in the experimentation.

This work was done under contract with the Bureau of Ordnance, United States Navy.

The Infra-Red Refractive Index and Dispersion of Evaporated Stibnite Films

BRUCE H. BILLINGS AND MARK HYMAN, JR.
Research Laboratories, Polaroid Corporation, Cambridge, Massachusetts

(Received October 28, 1946)

This paper describes two determinations which have been made of the refractive index at a wave-length of 10μ of a film of antimony trisulfide prepared by evaporation at a pressure of 10^{-4} mm of mercury. The first determination was made from transmission measurements of an antimony trisulfide film on a plate of silver chloride. The resulting index value was 2.43 at 10μ. The second determination was made from a dispersion curve which was extrapolated to a point in the visible spectrum where the index was obtained from a measurement of Brewster's angle. The dispersion curve was obtained from the values of the wave-lengths of maximum and minimum transmission of an antimony trisulfide film of optical thickness four waves at 632 mμ which had been prepared on a plate of sodium chloride. The resulting index value was 2.5 at a wave-length of 10μ.

IN the course of a study of stibnite-coated, high transmission silver chloride windows, as described in the preceding paper, it became desirable to measure the refractive index of an evaporated stibnite film at various wave-lengths in the infra-red.

The index of the antimony trisulfide layer was determined in two ways. In the first method, the index was derived from measurements of the transmission of a single layer on a silver chloride flat. The procedure is indicated in the following equations.

The general equation for the reflectivity R of one side of a material coated with a single transparent layer can be derived from electromagnetic theory and is

$$R=|\rho|^2, \tag{1}$$

where

$$\rho=\frac{r_{21}+r_{10}e^{i\theta}}{1+r_{10}r_{21}e^{i\theta}}, \tag{2}$$

where θ is the angular retardation for a double traversal of the coating layer; i is $(-1)^{\frac{1}{2}}$, r_{10} is the amplitude reflectivity of the interface between the coating and the base material, and r_{21} the amplitude reflectivity of the coating-air interface.

From Fresnel's laws for light at normal incidence, we may write for r_{10} and r_{21}

$$r_{10}=\frac{n_1-n_0}{n_1+n_0}, \quad r_{21}=\frac{n_2-n_1}{n_2+n_1}. \tag{3}$$

If n_1, the index of the coating layer, is greater than either n_0, the material index, or n_2, the index of the surrounding medium, and if these indices are all held constant, the condition for minimum transmission is $e^{i\theta}=-1$ and Eq. (2) becomes:

$$\rho=\frac{r_{21}-r_{10}}{1-r_{21}r_{10}}. \tag{4}$$

If we substitute into Eq. (4) the values for r_{10} and r_{21}, we may write

$$\rho=\frac{n_0n_2-n_1^2}{n_0n_2+n_1^2}. \tag{5}$$

The transmission through a thick plate of silver chloride with a stibnite coating can be

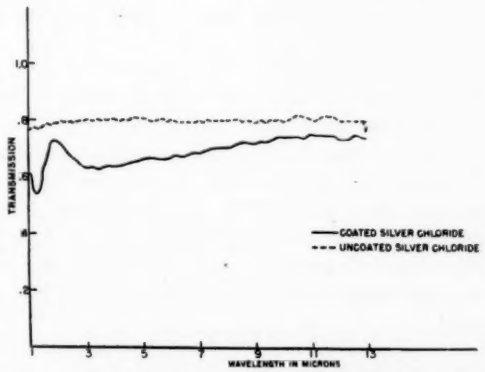

FIG. 1. Transmission of silver chloride flat coated on one side with an evaporated stibnite layer of optical thickness of 7/4λ at a wave-length of 625 mμ. The curve is irregular because it is drawn through a large number of experimental points.

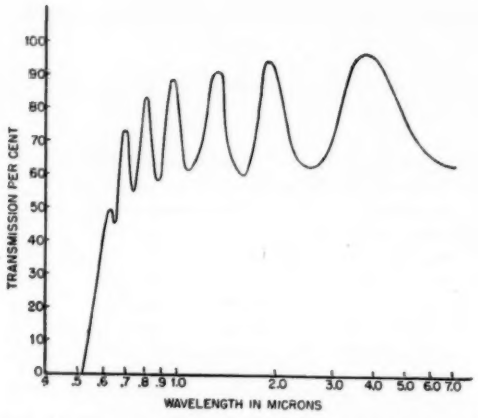

FIG. 2. Ratio of the transmission of a sodium chloride plate coated with evaporated stibnite to the transmission of an uncoated sodium chloride plate. The optical thickness of the Sb_2S_3 layer is four waves at a wave-length of 632 mμ.

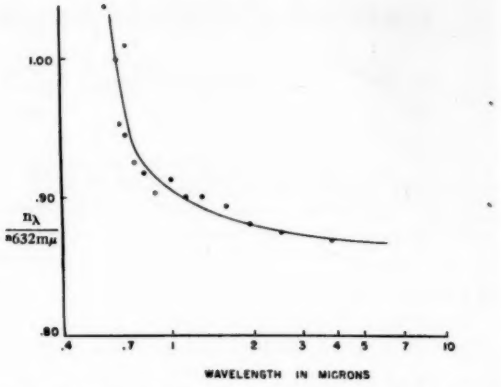

FIG. 3. Dispersion curve of evaporated stibnite. The index of refraction is measured with respect to the index at 632 mμ.

found from the following formula:

$$T=(1-R)(1-R_1), \tag{6}$$

where T is the transmission and R_1 is the reflectivity of the uncoated side, and where the multiple reflections have been assumed to combine incoherently. At 3.5μ the index of silver chloride is approximately 2.0; Eq. (6) thus becomes

$$T=0.89(1-R). \tag{7}$$

Figure 1 gives a typical transmission curve for a stibnite-coated silver chloride flat. The method used for forming such a coating is described in detail in the preceding paper; briefly, it involves the evaporation of stibnite from a crucible onto a silver chloride flat, the evaporation taking place at a pressure of 10^{-4} mm of mercury. The transmission maxima in this curve do not quite equal the values for uncoated silver chloride. This indicates that either during the coating process the silver chloride developed some slight degree of absorption or the Sb_2S_3 film is slightly absorbing. Considering this absorption factor, we must rewrite Eq. (8) as follows:

$$T=0.89k(1-R), \tag{8}$$

where k is the absorption.

From the 1.8μ maximum in Fig. 1, k is seen to be approximately 73/78. The absorption changes sufficiently slowly so that this same value of k may be used at 3.5μ, where the transmission is at a minimum and has a value of 0.63. If these values of k and T are inserted in Eq. (8), the solution for ρ is 0.494. With this value for ρ Eq. (5) may be solved and gives a value for the antimony trisulfide index of 2.43.

The second method depended on a measurement of the evaporated stibnite index at the sodium D line and a measurement of the dispersion of the evaporated layer. For the dispersion, a series of transmission measurements were made with a Beckman spectrometer in the visible and near infra-red regions of the spectrum and with a Perkin-Elmer spectrometer in the region of 1μ to 14μ. A curve is shown in Fig. 2 of such a set of transmission measurements made on a layer of antimony trisulfide evaporated on rocksalt. These measurements were made with respect to the transmission of a rocksalt plate without any evaporated layer. If the antimony trisulfide had been completely transparent, the maxima in this curve would lie along the hundred percent line. This can be demonstrated from Eq. (1). With $n_1>n_2$, n_0 the condition for maximum transmission is $e^{i\theta}=+1$. If we now substitute for r_{10} and r_{21} in Eq. (1) their values as given in Eq. (3), we find

$$\rho=(n_2-n_0)/(n_2+n_0). \tag{9}$$

This is the value we would get if the Sb_2S_3 film were absent. The spacing between the peaks in the transmission curve is a function of the dispersion. It was calculated from a number of

such samples using 632 mμ as a reference value. The dispersion curve is shown in Fig. 3.

The index at one wave-length in the visible was determined by Mr. Albert Makas of Polaroid Corporation. The Brewsterian angle method was used as developed by Pfund.[1] The index was measured on a rather thick film at the *D* line where there is high absorption so that reflection from the back surface would not cause difficulties. The value obtained was 3.01 ± 0.01. With this index value, plus the dispersion curve of Fig. 3 the index at 10μ can be calculated to be 2.5, in substantial agreement with the value obtained by the first method. These values are much lower than those known for the crystal stibnite, as shown by the following statement prepared by C. D. West:

"Although the refractive indices of stibnite as measured through prisms have been reported twice in the literature, the orientation of the indicatrix has remained in doubt in the latest reference works, the uncertainty arising from an inconsistency in the earlier citation. The following is believed to be a correct statement on the basis of published evidence:

	$n_{761}^{(3)}$	$n_{829}^{(3)}$	$n_{852}^{(3)}$
$X=b=$	3.194	—	—
$Y=a=$	4.046	3.919	3.875
$Z=c=$	4.303	3.875	4.137

The crystal is biaxial negative (−) with axial plane (100). The films described in the present account very probably have a non-crystalline structure; this could account, at least in part, for their lower refractive indices."

[1] A. H. Pfund, J. Opt. Soc. Am. **31**, 679 (1941).

[2] A. Hutchinson, Min. Mag. **14**, 199 (1907).

[3] R. Bailly, Ac. roy. Belgique Bull., Cl. Sci. **24**, 791 (1938).

Letter to the Editor

The Brightness of Apparent Fluorescence As a Function of the Exciting Intensities*

FRANZ URBACH,** ANNIE URBACH, AND MYRA SCHWARTZ
Institute of Optics, University of Rochester, Rochester, New York
January 6, 1947

IT is, on the whole, taken for granted that the brightness, or radiance, of fluorescence, or that of the stationary luminance observed during excitation, is strictly proportional to the intensity of the exciting light. However, some deviations from this simple relation have been found occasionally: a tendency toward saturation has been observed at extremely high intensities,[1] and in some zinc sulfide phosphors more than proportional dependence of the brightness on excitation has been reported at normal and elevated temperatures. With increasing intensities, increases of efficiency by a factor up to 1.7 were observed.[2]

We succeeded in obtaining with various phosphors much stronger deviations from linearity, first only at very low intensities, later at high intensities at elevated temperatures, and finally at high intensities also at room temperature. Changes of efficiency by a factor of several hundreds are observed if one or more of the following conditions are fulfilled to a sufficient degree:

1. low intensity of excitation,
2. high temperature during excitation,
3. presence in a phosphor of a so-called "poison" or, in some cases, of a second activator: the same effect is obtained by certain measures in the preparation which decrease the efficiency of the phosphor.

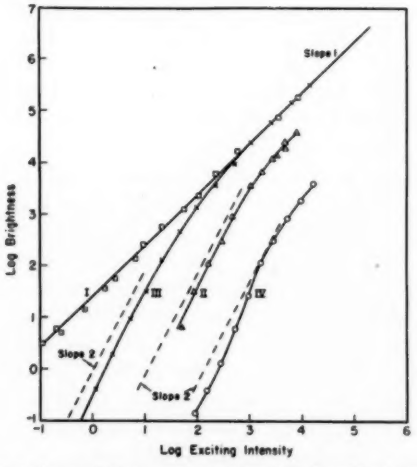

FIG. 1. Apparent fluorescence as a function of exciting intensities, log-log plots. I, linear fluorescence-ZnSCdS·Cu; II, non-linear fluorescence at high temperature—same phosphor as I at 185°C; III and IV, examples of non-linear fluorescence at room temperature—ZnS·AgNi phosphors. (The units are arbitrary and different for different curves.)

The intensity, temperature, and the amounts of poisons needed for producing strong effects are interdependent. Thus, the well-known drop of efficiency at high temperatures depends strongly upon the intensity level at which it is observed. Low intensities, for instance, shift that drop to lower temperatures. Accordingly, data about the temperature dependence of the luminescence of many phosphors are significant only if the intensity of excitation is taken into account.

In a similar manner, the effect of a poison may be slight at high intensities and strong at low intensities, a fact that explains some apparently contradictory observations on poisoning effects. Furthermore, the difference between excitation by radiations of different wave-lengths, by fast electrons, and by x-rays should be at least partly dependent upon the density of excitation produced by the different radiations. Also, the selective reduction of afterglow by poisoning or by increasing temperature is probably partly determined by the dependence of efficiency upon the density of excitation.

As one general scheme of interpretation, we have considered the possibility that the two processes of spontaneous decay of excitation, radiative and non-radiative, may be of different orders of reaction. The place of the non-radiative processes might be taken by radiative processes, provided their emission is not observed or is not observable under the conditions of the experiment. If the number of the excited states is denoted by n, the equilibrium during excitation is given by

$$dn/dt = aI - bn^2 - cn = 0, \qquad (1)$$

where a, b, and c, are constants. If the brightness, B, is given by the second term only, $B = bn^2$, one obtains at sufficiently high intensity a proportionality of B with I. If the intensity is sufficiently low, B becomes proportional to the square of the exciting intensity.

A generalization of this scheme, and models implementing it, will be discussed elsewhere. Some measurements seem to agree quite well with such an interpretation, as Fig. 1 curves II and III show. However, cases have been observed in which the brightness increases even more rapidly than with the second power of the exciting intensity (Fig. 1, curve IV). This will require at least a modification of Eq. (1).

* This work was done in connection with a NDRC project, mostly during the first few months of 1943. Some results were mentioned and demonstrated at the meeting of the Optical Society of America, in Cleveland, March, 1946. The declassification of the work on the "Icaroscope" made it possible to present a few more details at the Optical Society meeting in New York, September, 1946.
** Now in the Research Laboratories, Eastman Kodak Company, Rochester, New York.
[1] W. de Groot, Physica 6, 393–400 (1939).
[2] N. Riehl, Zeits. f. tech. Physik 20, 152–55 (1939). J. H. Gisolf and F. A. Kröger, Physica 6, 1101–11 (1939).